INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Leitura automática de números

Maria Cristina Loureiro Lopes

Relatório Final da Dissertação apresentado à
Escola Superior de Tecnologia e de Gestão
Instituto Politécnico de Bragança

para obtenção do grau de Mestre em
Engenharia Industrial

Janeiro de 2015

INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Leitura automática de números

Maria Cristina Loureiro Lopes

Relatório Final da dissertação apresentado à
Escola Superior de Tecnologia e de Gestão
Instituto Politécnico de Bragança

para obtenção do grau de Mestre em
Engenharia Industrial

Orientador:
Prof. Doutor João Paulo Teixeira

Este Trabalho de Dissertação
"inclui as críticas e sugestões feitas pelo Júri"

Janeiro de 2015

Agradecimentos

Começo por agradecer de forma muito especial aos meus pais, irmãos e avós, que sempre me apoiaram ao longo desta caminhada.

Ao meu orientador desta tese, Professor Dr. João Paulo Teixeira, pela oportunidade de realização deste trabalho, pelo seu apoio e orientação durante a minha vida académica, pelo incentivo, força, compreensão e ajuda demonstrada ao longo da realização deste trabalho.

Ao Sr. Paulo Afonso, da Rádio Brigantia, pela cedência das instalações e condições técnicas para a recolha dos sinais de fala.

A todos os meus amigos, pela força e ajuda oferecida em todos os momentos, fazendo assim com que esta batalha fosse vencida.

Por último mas não menos importante quero agradecer de forma muito especial à minha filha Beatriz, que mesmo sem saber foi a pessoa que mais motivação me deu. Também ao meu marido Luís pelo carinho, força, apoio e compreensão que me dedicou fazendo com que este trabalho fosse possível.

Resumo

Neste trabalho foi desenvolvido um sistema que faz a leitura automática de números. A entoação dada a um número em posição final de uma sequência é diferente da entoação usada para números noutras posições.

O sistema faz a leitura, dos números inteiros em Português Europeu, de 0 (zero) a 999 999 999, (novecentos e noventa e nove milhões, novecentos e noventa e nove mil, novecentos e noventa e nove), de datas no formato (dd-mm-aaaa), números de telefone da rede fixa, números de telemóvel e número de identificação da segurança social.

O sistema começa por identificar o tipo de número, depois faz a leitura desse número usando o algoritmo desenvolvido para cada caso. Estes foram programados utilizando o software Matlab.

Os sinais de áudio foram gravados na "Radio Brigantia", com a voz de um locutor profissional, do sexo masculino e editados utilizando o software Praat.

Finalmente foi realizado um teste auditivo para avaliar a qualidade dos sons para cada algoritmo. Cada um deles foi avaliado numa escala MOS (Mean Opinion Score) de 1 a 5. A pontuação MOS do trabalho desenvolvido foi de 4,46.

Palavras-chave: leitura automática de números, sistemas TTS, concatenação de números, leitura de números em português, fala.

Abstract

This work presents the development of a system that makes the automatic reading of numbers. The prosody given to a number in a final position is different from the one used for numbers elsewhere.

The system reads, the integers in European Portuguese, from 0 (zero) to 999,999,999 (nine hundred ninety-nine million, nine hundred ninety-nine thousand, nine hundred ninety-nine), dates in the format (dd-mm-yyyy), wireline phone numbers, cell phone numbers and social security number of identification.

The system begins by identifying the type of number, then reads that number using the algorithm developed for each case. These were programmed using Matlab software.

Audio signals were recorded in the "Radio Brigantia" with the voice of a professional announcer, male and edited using the Praat software.

Finally we conducted a hearing test to assess the quality of the sounds for each algorithm. Each was evaluated on a MOS (Mean Opinion Score) scale 1 to 5. The work MOS score was 4.46.

Keywords: Automatic reading numbers, TTS systems, concatenation number, reading numbers in Portuguese, speech.

Índice

Índice de figuras

Índice de Tabelas

1 – Introdução

1.1 – Objetivos

O objetivo deste trabalho foi desenvolver um sistema de leitura automática que permita a leitura de números entre 0 (zero) e 999 999 999 (novecentos e noventa e nove milhões, novecentos e noventa e nove mil, novecentos e noventa e nove), números de telemóvel, números de telefone, número de identificação da segurança social e datas no formato (dd-mm-aaaa).

Foram desenvolvidos cinco algoritmos diferentes, cada um com a sua própria estrutura, funções e características. A síntese utilizada para este fim específico consiste na concatenação de segmentos gravados.

Basicamente, o atual sistema automático de leitura de números, começa por ler os números introduzidos pelo utilizador e procede à sua identificação ou seja a qual grupo o número pertence, unidades, dezenas, centenas, milhares, dezenas de milhar, centenas de milhar, milhões, dezenas de milhões e centenas de milhões. Depois disso, é necessário fazer a concatenação dos sons da fala correspondentes aos números. Se for necessário, a partícula "e" é adicionada, como por exemplo no número 1100 (mil **e** cem), 1200 (mil **e** duzentos), 1500 (mil **e** quinhentos), etc.

1.2 - Enquadramento do trabalho

Desde sempre, que o ser humano necessita de comunicar e de se expressar através da linguagem. É uma qualidade que pode ser expressa de várias formas [1]. No caso específico deste trabalho é a quantidade de algo expresso por números. No dia-a-dia os números são usados para quase tudo e por todos, tais como: uma data, um número de telefone, a idade, hora, uma quantidade monetária, um código postal, o número de cartões de identificação são alguns dos exemplos comuns.

É apresentada uma breve introdução sobre sistemas texto-fala, a sua estrutura principal e alternativas. Normalmente, o número aparece dentro de um texto como uma sequência de números que devem ser convertidos num sistema Texto-Fala. A forma como é feita a conversão do número depende daquilo que ele representa. Por exemplo, a conversão de um número de telefone, vai ser diferente da conversão de uma data, ou de uma quantidade de dinheiro. No caso de o número de telefone é mais apropriado converter dígito a dígito, a data vai ser convertida em dia/mês/ano, enquanto que a quantidade deve ser convertida num número inteiro. Além disso, a prosódia deve ser apropriada para cada caso. O número de telefone é lido como uma sequência de grupos de dígitos.

Por esta razão, cada tipo de número deve ser identificado para ser lido e convertido corretamente. Existem alguns aspetos importantes a ter em consideração, na conversão de números, tais como:

1º Identificação do tipo de número;

2º A conversão dos caracteres numéricos para texto ou alguma representação simbólica do texto correspondente;

3ºSíntese do texto correspondente.

É por esse motivo que, é tão importante ter algum sistema ou programa que possa ler os números automaticamente, para serem usados em algumas aplicações. Por exemplo, uma interface com um PC usando a interface de voz é muito útil para a pessoa cega ou uma interação surdo-mudo, mas também para várias outras aplicações.

O trabalho apresentado foi implementado na plataforma de desenvolvimento do Matlab.

1.3 – Sistema de produção da fala

O sistema de produção da fala está intrinsecamente ligado aos órgãos e sistema de respiração. No processo de expiração são permitidas maiores variações de pressão do que no processo de inspiração, tornando-se audível, pela produção de ondas sonoras que, modeladas pela laringe e as cavidades superiores orais e nasais, darão as características da voz.

No ato da expiração o fluxo de ar segue um trajeto inverso ao trajeto seguido pelo ar no processo de inspiração, saindo dos alvéolos, passando pelos brônquios, traqueia, laringe, faringe e finalmente cavidade nasal e/ou oral, como mostra a figura 1.

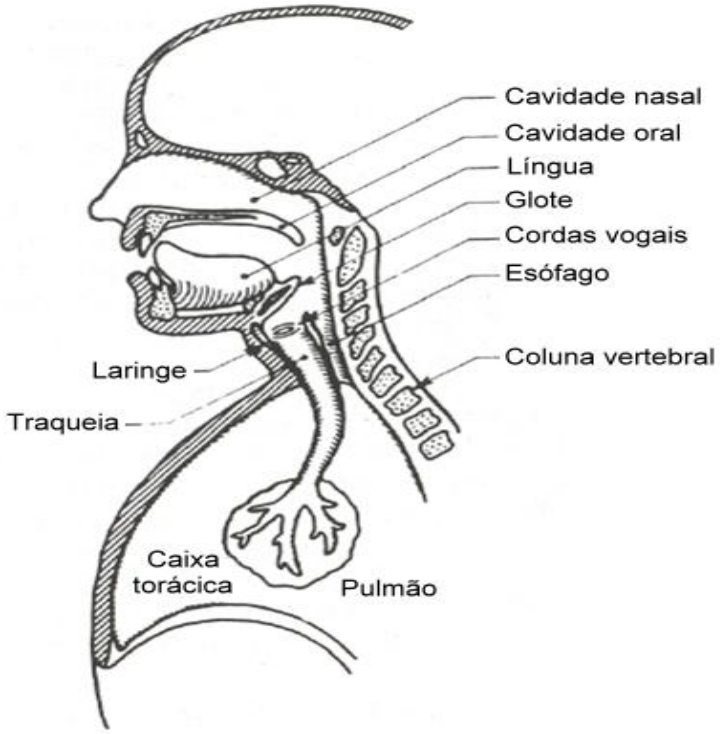

Fig. 1 – Vista esquemática do mecanismo vocal humano [2]

Logo que o fluxo de ar saia dos pulmões (alvéolos e brônquios) e depois da passagem pela traqueia, passa pela laringe que desempenha um papel fundamental no processo de produção da fala. É no interior da laringe que se situam as cordas vocais, compostas por ligamentos e músculos.

Na parte superior situam-se as cordas ventriculares (glandulares). Na parte inferior situam-se as cordas vocais, que são musculares. O espaço entre as duas cordas vocais, direita e esquerda é a glote [2], como mostra a figura 2.

Fig. 2 – Movimento vibratório das cordas vocais [2]

Durante a fonação, a função das cordas vocais é atuar como um gerador de som, abrindo e fechando rapidamente a passagem ao fluxo de ar vindo dos pulmões. A junção das cordas vocais cria pressão subglotal que vai aumentando até ser suficiente para afastar as cordas vocais uma da outra. Quando as cordas vocais se afastam, o ar sai, a pressão decresce, e as cordas voltam a aproximar-se. A cada ciclo destes dá-se o nome de período fundamental e gera um impulso glotal. A cadência de impulsos glotais por segundo é a frequência fundamental (F0). Este ciclo repete-se durante a fonação.

Continuando o trajeto do ar expirado, segue-se a faringe que se situa entre a laringe e as cavidades oral e nasais.

As cavidades nasais são duas e estão separadas pelo septrum nasal. A base ou "chão" das cavidades nasais é constituído pelo palato duro.

A cavidade oral situa-se entre a faringe posterior e os dentes. A parte inferior é constituída pelo "chão" da boca e pela língua e na parte superior pelo palato duro. A zona interna da implantação dos dentes é a zona dos alvéolos ou alveolar. Os dentes do maxilar superior estão fixos enquanto que os do maxilar inferior acompanham os movimentos deste. A parte oral do palato estende-se entre os alvéolos dos dentes e a úvula. Depois dos alvéolos encontra-se uma zona de "rugas" – o palato duro – com um silhão palatino, depois o véu palatino – ou palato mole – e a úvula. A língua é um órgão muscular coberto por uma membrana mucosa. Os bordos anteriores e laterais estão soltos; a parte inferior está ligada ao "chão" da boca e ao osso hióide.

Na parte posterior-inferior está ligada a epiglote. A massa principal da língua é constituída por uma série de quatro músculos: o superior longitudinal, o inferior longitudinal, o vertical e o transversal que permitem, durante a elocução, o alongamento ou constrição da língua no seu todo, ou só em segmentos específicos necessários à articulação dos sons da fala [3].

A síntese da fala é a produção artificial da voz humana. Um sistema de computador utilizado para este efeito é o chamado sintetizador de fala, e pode ser implementado em software ou hardware. O sistema texto-fala (TTS), converte texto em linguagem normal de fala, outros sistemas tornam representações linguísticas simbólicas como transcrições fonéticas em discurso.

Basicamente, fala sintetizada pode ser criada concatenando segmentos de voz gravada sintetizada que é armazenada numa base de dados. Os sistemas diferem no tamanho das unidades do segmento de voz armazenado. Para domínios de utilizadores específicos, o armazenamento de palavras ou frases inteiras permite a saída de alta qualidade. Alternativamente, um sintetizador pode incorporar um modelo do trato vocal e outras características da voz humana para criar uma saída de voz completamente sintética.

A qualidade de um sintetizador de voz é avaliada pela sua semelhança com a voz humana e pela sua capacidade de ser entendida. Um programa inteligível texto-fala, permite que pessoas com deficiência visual ou dificuldade de leitura possam ouvir textos escritos na sua própria casa. Muitos computadores têm incluído sintetizadores de voz no seu sistema operativo [4].

1.4 – Descrição da lista de fonemas do Português

Um fonema é a menor unidade sonora de uma língua que estabelece contraste de significado para distinguir palavras. Os fonemas não podem ser confundidos com letras. Enquanto o fonema é o som em si mesmo, a letra é a representação gráfica desse som. É bastante comum que um mesmo fonema seja representado por diferentes letras, também pode acontecer a mesma letra representar mais de um fonema.

1.4.1 – Caraterização de alguns traços fonéticos

Irei caracterizar cada um dos traços distintivos acústicos do ponto de vista acústico e articulatório, bem como a caracterizar os fonemas ou classe de fonemas com estes traços, uma vez que são definidos pelas suas características espectrais.

1.4.1.1 – Traços de sonoridade

Na tabela I apresentam-se os traços de sonoridade com base nas referências [5], [6], [7].

Tabela I – Traços de sonoridade [5], [6], [7]

Traços de sonoridade	**Caracterização acústica**	**Caracterização articulatória**	**Exemplos**
- Vocálico/Não vocálico	- Presença/Ausência de uma estrutura formântica. - Espectro de energia decrescente ao longo das frequências. - Sons vocálicos apresentam uma energia superior aos sons não vocálicos.	- Excitação da glote associada a uma passagem livre do ar através do trato vocálico.	- Vocálico: vogais e liquidas. - Não vocálico: deslizantes e obstruentes.

- Consonântico/ não consonântico	- Presença de uma energia globalmente baixa em oposição à energia globalmente alta. - Presença de anti-ressonâncias que afetam todo o espectro. - Larguras de banda significativamente maiores.	São produzidos por uma obstrução significativa do trato vocálico.	Consonântico: consoantes obstruentes e líquidas. Não consonântico: vogais e deslizantes.
- Compacto/Difuso	- Predominância de formantes na zona central do espectro em oposição à predominância de formantes fora da zona central do espectro.	- São produzidos com uma configuração do trato vocal em que o volume das cavidades são próximos.	- Compacto: vogais abertas, consoantes velares e palatais. - Difuso: vogais fechadas, consoantes alveolares, dentais e labiais.
- Tenso/Relaxado	- Um som tenso tem maior duração e energia global.	- Maior afastamento da configuração do trato vocálico corresponde uma maior tensão muscular e um maior aumento da pressão.	- Tenso: vogais e consoantes longas e consoantes aspiradas. - Relaxado: vogais e consoantes breves e consoantes não aspiradas.

- Vozeado (ou vocalizado) / não vozeado (ou não vocalizado)	- Presença/Ausência de uma excitação periódica de baixa frequência. - Estrutura formântica nítida/ou não.	- Vibração periódica das cordas vocais.	- Vozeado: vogais, liquidas, deslizantes, nasais e consoantes [b,d,g,v,z,j]. - Não vozeado: consoantes [p,t,k,f,s].
- Nasal/Oral	- Introdução de anti-ressonâncias e frequências de ressonância adicionais. - Aumento da largura de banda. - Diminuição da energia dos formantes, particularmente de F1.	- Produzido pela ressonância da cavidade nasal associada à ressonância da cavidade oral.	- Nasal: consoantes [m, n, nh] e vogais nasais. - Oral: todas as que não são nasais.
- Contínuo/Não contínuo	- Nos sons não contínuos ou interrompidos existe um ataque abrupto associado a alterações bruscas nas características espectrais em oposição aos sons contínuos em que o ataque é gradual.	- Os sons não contínuos são caracterizados pelo surgimento ou desaparecimento rápido da fonte, devido a uma oclusão ou a uma abertura do trato vocálico.	- Contínuo: fricativas e laterais. - Não contínuo: oclusivas e vibrantes.

- Estridente/não Estridente	- Maior intensidade de ruído. - Apresentam formas de onda extremamente irregulares, apresentando uma distribuição aleatória das energias no espectro.	- O fluxo de ar é dirigido contra um obstáculo na vizinhança da constrição dando origem a um forte ruído de turbulência.	- Estridente: fricativas [f, s, x, u, z, j].
- Bloqueado/não Bloqueado	- Terminação abrupta que é no entanto menos proeminente que um ataque abrupto.	- Produzido por uma compressão ou oclusão total da glote.	- Bloqueado: implosivas, cliques e ejectivas.

1.4.1.2 – Traços de tonalidade

Na tabela II apresentam-se os traços de tonalidade conforme referência [5].

Tabela II – Traços de tonalidade [5]

Traços de tonalidade	**Caracterização acústica**	**Caracterização articulatória**	**Exemplos**
- Grave/Agudo	- Concentração da energia numa zona baixa das frequências no espectro em oposição ao traço agudo a que corresponde uma concentração de energia numa zona alta de frequências no espectro	- Para o traço grave a cavidade de ressonância é mais ampla e menos compartimentada, pelo que este traço se produz em zonas periféricas em oposição ao traço agudo que se produz em zonas centrais.	- Grave: vogais graves (exemplo [u]) e consoantes labiais e velares. - Agudo: vogais agudas (exemplo [i]) e consoantes dentais e palatais
- Bemolizado/não Bemolizado	- Redução da energia de alguns ou todos os formantes.	- Produz-se por um arredondamento dos lábios e aumento da cavidade anterior à constrição.	- Bemolizado: consoantes labiolizadas e as vogais [o, u]
Diesado/não Diesado	- Elevação das frequências nas componentes de frequências mais altas.	- Elevação do corpo da língua com uma dilatação da cavidade faríngea.	- Permite distinguir consoantes palatizadas de não palatizadas.

1.4.2 – Parâmetros classificatórios das consoantes

Existem dois grandes parâmetros classificatórios específicos das consoantes: o modo de articulação e o ponto de articulação. Correspondendo ao primeiro o modo da passagem do ar pelo trato vocal e o segundo à região do trato vocal em que se situa a maior constrição imposta pelos articuladores no canal bucal.

1.4.2.1 – Modo de articulação

A sua classificação é função da aproximação dos articuladores, da duração dessa aproximação, ou da modificação da configuração do trato vocal devido à aproximação dos articuladores superiores e inferiores.

1.4.2.1.1 – Consoantes Oclusivas

A articulação das consoantes oclusivas é realizada pelo impedimento da passagem de ar pelo canal bucal por um fechamento completo dos articuladores, ou seja são as consoantes pronunciadas fechando-se totalmente o aparelho fonador, sem dar espaço para o ar sair. A tabela III apresenta a lista das oclusivas orais para o Português Europeu.

Tabela III – Oclusivas orais para o Português [5]

Oclusivas		Causa da oclusão
Vozeada	Não vozeada	
[b]	[p]	Fechamento dos lábios.
[d]	[t]	Coroa da língua encostada aos incisivos superiores.
[g]	[k]	Dorso da língua encostada ao véu palatino.

Caso o fluxo de ar puder passar pelas cavidades nasais, devido ao véu palatino estar descido, produz-se uma oclusiva nasal. Estas consoantes são sempre vozeadas. A tabela IV apresenta a lista das oclusivas nasais para o Português Europeu.

Tabela IV – Oclusivas nasais para o Português [5]

Oclusiva nasal	**Causa da oclusão**
[m]	Fechamento dos lábios.
[n]	Coroa da língua encostada aos incisivos superiores.
[nh]	Lâmina da língua encostada ao palato.

1.4.2.1.2 – Consoantes fricativas

São as consoantes pronunciadas através de uma corrente de ar que se fricciona em um obstáculo. A tabela V apresenta as 6 fricativas para o Português Europeu.

Tabela V – Fricativas para o Português [5]

Fricativa		Modo de obstrução à passagem de ar
Vozeada	Não vozeada	
[v]	[f]	O lábio superior aproxima-se dos incisivos inferiores
[z]	[s]	A coroa da língua aproxima-se da região dento-alvelar.
[j]	[x]	A coroa da língua aproxima-se da região palato-alvelar.palatino.

1.4.2.1.3 – Consoantes laterais

São as consoantes pronunciadas ao fazer passar a corrente de ar nos dois cantos da boca ao lado da língua. A consoante lateral [l] pode ser velar, quando surge precedida de consoante ou em posição final (ex. a**l**to, ma**l**). A tabela VI apresenta as duas consoantes laterais do Português Europeu.

Tabela VI – Laterais para o Português [5]

Consoante lateral	**Obstrução**
[l]	Formada pela ponta da língua junto dos alvéolos
[lh]	Formada pela lâmina da língua junto ao palato

1.4.2.1.4 – Consoantes vibrantes

São as consoantes pronunciadas através da vibração de algum elemento do aparelho fonador, em geral a língua ou o véu palatino. A tabela VII apresenta as consoantes vibrantes do Português Europeu.

Tabela VII – Vibrantes para o Português [5]

Consoante vibrante	**Denominação**	**Articulação**
[r]	Vibrante alveolar	Uma única obstrução provocada pela ponta da língua junto dos alvéolos, (exemplo: caro).
~ [r]	Vibrante alveolar múltipla	A ponta da língua toca várias vezes nos alvéolos, (exemplo: carro).
[R]	Vibrante velar	Vibração da parte de trás da língua junto do velo.

1.4.2.1.5 – Consoantes africadas

Iniciam-se por uma oclusão completa e terminam com uma constrição própria das fricativas. Em Portugal a única consoante africada é não vozeada e existe apenas em alguns dialetos, representa-se por [tx] (exemplo: tchau). No Brasil as africadas podem ser vozeadas, [dz], ou não, [tx].

1.4.2.2 – Ponto de articulação

Na tabela VIII apresentam-se as consoantes classificadas quanto ao seu ponto de articulação.

Tabela VIII – Ponto de articulação [5]

Denominação quanto ao ponto de articulação	**Articuladores**	**Exemplos**
Bilabiais ou labiais	Os dois lábios.	[b], [p], [m]
Labiodentais	O lábio inferior e os incisivos.	[v], [f]
Dentais	Ponta da língua e os incisivos.	[d], [t], [z], [s], às vezes [e], [n]
Alveolares	Ponta da língua e os incisivos e os incisivos superiores.	[l], [n], [r]
Pré-palatais	A lâmina da língua e o pré-palato.	[z], [x]
Platais	A lâmina da língua e o palato.	[lh], [nh]
Velares	A parte de trás da língua e o véu palatino.	[g], [k], [R]

1.4.3 – Lista de fonemas do Português

De seguida é apresentado o alfabeto fonético para o Português Europeu baseado em [5], bem como a sua representação fonética usando o Alfabeto fonético Internacional e um exemplo de cada fonema numa palavra. Quando uma vogal aparece junto de uma semi-vogal, chama-se ditongo ao conjunto das duas.

Vogais orais		**Vogais nasais**		**Semi-vogais**	
[i]	l**i**vro	[ĩ]	p**in**to	[j]	pa**i**
[e]	P**e**dro	[ẽ]	d**en**te	[w]	pa**u**
[ɛ]	t**e**rra	[ɑ̃]	c**an**to		
[a]	p**a**to	[õ]	p**on**te		
[ɑ]	m**a**no	[ũ]	f**un**do		
[ɔ]	g**o**la				
[u]	p**u**lar				
[ə]	se**c**ar				
[o]	p**o**ço				

Consoantes oclusivas orais		**Consoantes oclusivas nasais**	
[p]	**p**ara	[m]	a**m**a
[b]	**b**ata	[n]	**n**ada
[t]	**t**arde	[ɲ]	pi**nh**o
[d]	**d**a**d**o		
[k]	**c**ão		
[g]	**g**ato		

Consoantes fricativas		**Consoantes laterais**		**Consoantes vibrantes**	
				~	
[f]	**f**ado	[l]	**l**ado	[r]	po**r**ta
[s]	**s**ábado	[ʎ]	fi**lh**o	[r]	ca**rr**o
[ʃ]	**ch**ão				
[v]	**v**aca				
[z]	ca**s**a				
[ʒ]	**j**ardim				

2 - Conversores Texto-Fala

Pretende-se apresentar sumariamente alguns conceitos básicos sobre síntese de fala e conversores texto-fala, tais como: conversores texto-fala ou aplicações de reconhecimento de fala.

2.1 - Síntese da Fala e Conversores Texto-Fala

As primeiras máquinas de produção de fala, muito primitivas, remontam a 1779, por C.G. Kratezenstein. Alguns anos mais tarde, em 1791, W.R. von Kempelen demonstrou uma máquina mais sofisticada, capaz de reproduzir fala contínua. Em 1835, Wheatstone melhorou a máquina de von Kempelen e construiu um dos primeiros *vocoders*, ainda manual e analógico, como se pode ver na figura 3.

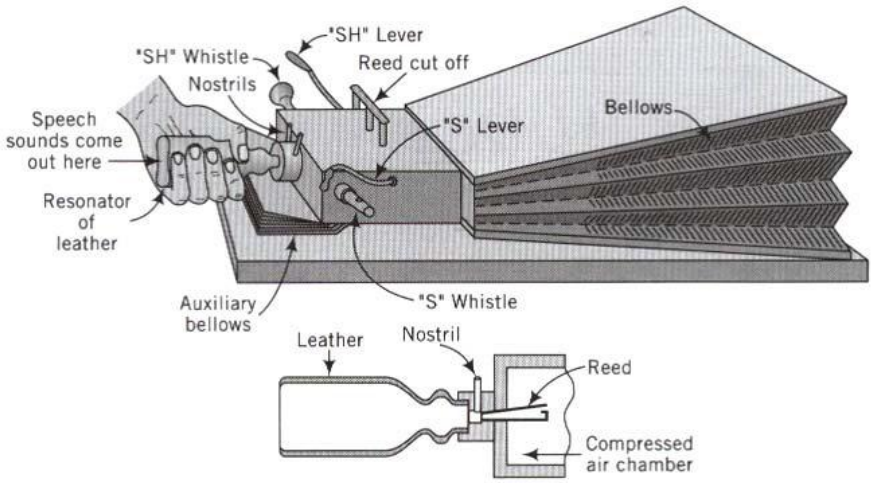

Fig. 3 - Máquina Falante de Wheatstone [1]

Em 1939, Homer Dudley inventa o primeiro vocoder elétrico e demonstra-o na feira mundial em Nova York. Ainda em 1939, Dudley propõe que os *vocoders* passem a representar os sinais de fala no domínio da frequência modelando-os através de um conjunto de filtros passa-banda. Este conjunto de filtros é excitado por ruído nas zonas não vozeadas (zonas em que as cordas vocais não vibram) e por pulsos periódicos nas zonas vozeadas (zonas em que o ar liberto pelos pulmões é colocado em vibração pela

ação das cordas vocais). Atualmente, ainda existem muitos sistemas que fazem uso deste princípio básico [1].

Com o aparecimento dos computadores, o processamento da fala desenvolveu-se rapidamente e surgiram diversas tecnologias ao serviço da engenharia da fala.

O desenvolvimento das técnicas de síntese de fala, conjugadas com a evolução dos sistemas informáticos, conduziu ao aparecimento dos conversores texto-fala. Poder-se-ia definir um conversor texto-fala como um sistema baseado num computador capaz de processar um texto e reproduzi-lo auditivamente.

2.2 - Sistema Texto-Fala

Um sistema texto-fala ou TTS, converte texto em linguagem normal para voz. A voz sintetizada pode ser criada concatenando-se pedaços de fala gravada, armazenada numa base de dados. Os sistemas diferem no tamanho das unidades de fala armazenadas, um sistema que armazena fones (segmento vocálico), ou alofones (variedade fonética), fornece maiores parâmetros de saída, mas pode ser necessário maior clareza. Para usos específicos, o armazenamento de palavras ou frases inteiras possibilita uma saída de alta qualidade. Alternativamente, um sintetizador pode incorporar um modelo de trato vocal e outras características da voz humana, para criar como saída uma voz completamente sintética.

A qualidade de um sintetizador de voz é determinada, primeiro pela sua inteligibilidade e depois pela similaridade com a voz humana. Atualmente existe uma grande variedade deste tipo de sistemas, aplicações de negócio de desenvolvimento para fins educacionais. Um programa TTS inteligível para pessoas com deficiência visual, ou a nível da fala, permite que o sintetizador de fala associado a um PC se faça ouvir com recurso a um programa específico de onde se podem selecionar e compor rapidamente e com facilidade um grande número de mensagens pré-gravadas ou escritas no momento, podendo assim estabelecer comunicação.

Um sistema TTS completo é dividido em duas partes, o processamento linguístico-prosódico e o processamento acústico [5], conforme representado na figura 4.

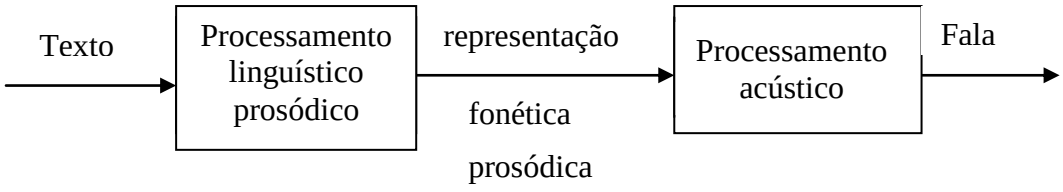

Fig.4 – Diagrama de blocos de um sistema de conversão texto fala

O texto é a entrada de um sistema TTS, em seguida, no módulo vão ocorrer várias tarefas linguístico-prosódias, como o pré-processamento de texto utilizado para a conversão de números, abreviaturas, siglas e outros caracteres em texto para ser processado pelo sub-módulo linguístico.

O sub-modelo linguístico converte morfemas ou texto em fonemas no nível segmental, identifica vários excertos de informação tais como: frases, números, palavras, sílabas e fonemas. Alguns módulos linguísticos avançados também realizam uma análise sintática e / ou gramatical se o analisador correspondente estiver disponível.

É objetivo do processamento linguístico-prosódico determinar, a partir de um texto, dois tipos de informação necessária para proporcionar ao processamento acústico dados que lhe permitam gerar uma fala o mais natural possível. Estes dois tipos de informação são conhecidos como informação segmental e informação suprassegmental.

A informação segmental está associada aos sons elementares que compõem a mensagem. Para cada língua existe um conjunto limitado de sons base ideais que permitem reproduzir, quando corretamente combinados, todas as particularidades da fala nessa língua. Criam-se assim uma série de representações abstratas denominadas fonemas cuja variedade depende da língua em causa.

A informação suprassegmental está associada à prosódia, reflete tanto elementos linguísticos (tais como tipo de frases, pausas acentuação e agrupamento de elementos com algum significado), como elementos não linguísticos. Esta informação é considerada por muitos autores a chave para conseguir uma elevada naturalidade na fala sintetizada [5]. A informação suprassegmental vem geralmente codificada através de três parâmetros acústicos do sinal de fala:

a). A evolução temporal da frequência fundamental, que é o aspeto mais importante do ponto de vista percetivo;

b). Duração dos segmentos de som que compõem a frase;

c). Curva de energia do sinal acústico.

A identificação e modulação da prosódia é uma tarefa difícil. Para modelar as durações segmentais vários modelos têm sido utilizados, tais como o modelo Z-score [9], modelo Barbosa & Baailly [10], ou modelos baseados em redes neurais artificiais [11]. F0 é o parâmetro acústico mais importante para transmitir a prosódia. Vários modelos/ técnicas têm sido estudadas para modelar estes parâmetros como o modelo de Fujisaki [11] e [12], o modelo ToBI (Tone and Break Indices) [13], o modelo de TILT [14] ou o INTSINT [15].

O módulo de processamento acústico produz o sinal acústico da fala correspondente à sequência de fonemas e com a prosódia modelada em blocos de processamento anteriores. Diversos métodos foram sendo usados desde o modelo de formantes Klatt [16], os modelos lineares LPC tendo sido o modelo RELP (Prediction Residual Excitação Linear) o mais aceitável em termos de qualidade, mas já não está em uso, o modelo sinusoidal [17] também obsoleto. Os modelos mais comuns têm sido o PSOLA e o Time Domain PSOLA [18], um modelo de concatenação ainda com boa qualidade, atualmente. Os modelos articulatórios [19] que não estão em uso, modelos de seleção das unidades [20], modelo com muito boa qualidade, e os modelos baseados em HMM [21] que foram muito promissores.

As qualidades mais importantes de um sistema de síntese de fala são a naturalidade e inteligibilidade. Naturalidade descreve quão próximo a saída dos sons se assemelham à fala humana, enquanto que a inteligibilidade é a facilidade com que a saída é compreendida. O sintetizador de fala ideal é natural e inteligível. Os sistemas de síntese de fala geralmente tentam maximizar ambas as características.

2.3 - Aplicações de um conversor Texto-Fala

São inúmeras as aplicações de um sistema texto-fala, tais como:

- em serviços de telecomunicações, uma vez que muitas empresas fazem uso de conversores texto-fala nos centros de atendimento automático (call center), ou para serviços especializados;
- na formação e ensino, em que alguns conversores são utilizados na aprendizagem de novas línguas;
- em multimédia, o uso de conversores texto-fala tem aberto novas fronteiras na interface pessoa-máquina;
- na diversão e comércio, existem produtos no mercado de brinquedos e livros que "falam";
- em investigação laboratorial, um conversor texto-fala é uma ferramenta linguística que pode ser controlada e modificada para a investigação de novas teorias e conceitos nas tecnologias de fala;
- na acessibilidade, inclusão e terapêutica da fala – pessoas com necessidades especiais, nomeadamente cegos e amblíopes, deficientes motores, entre outros, fazem uso de conversores texto-fala. A síntese da fala serve como ferramenta para a terapia vocal, e como ferramenta de acessibilidade.

Tal como os conversores Texto-Fala, também um sistema de leitura automática de números tem diversas aplicações, tais como: em interfaces homem-máquina, em pessoas com necessidades especiais, nas telecomunicações, em aplicações onde a variedade de testes a ser produzido é limitado a um determinado domínio conhecido previamente, como anúncios de agendamento de trânsito ou até mesmo previsões de tempo, em dispositivos de voz, etc.

3 – Construção e tratamento da base de dados

A síntese do número identificado consiste na simples concatenação dos dígitos e das partículas correspondentes, retirados de um contexto semelhante.

A base de dados dos sons consiste no segmento de fala de cada dígito e partículas em diferentes posições.

Os sons foram gravados, no estúdio da Rádio Brigantia, pelo locutor profissional Sr. Paulo Afonso, com todos os cuidados inerentes a um estúdio de gravação. A gravação foi feita com uma resolução de 16 bits e uma frequência de amostragem de 44100 Hz, em stereo, tendo sido posteriormente todos os sons tratados, através da função decimate do matlab para reduzir a frequência de amostragem para 22050 Hz, para evitar que a base de dados se tornasse demasiado pesada e o programa fosse mais rápido a reproduzir os sons, adicionalmente todas as gravações foram convertidas em mono.

A base de dados é constituída por duzentos e oitenta segmentos de voz gravados e ocupa no disco cerca de 6,1 mega bytes (MB). Em anexo, encontra-se a lista dos segmentos de voz gravados na base de dados e a lista dos sons gravados pelo locutor.

Durante o processo de gravação, tentou-se garantir que o locutor mantivesse o tom de voz, os mesmos números foram gravados várias vezes em diferentes posições e para que fosse possível obter um número em posição final, o locutor teve que fazer uma pausa de pelo menos 10 segundos entre cada gravação, para que o número final tivesse a entoação própria de um dígito lido numa posição final.

Para construir a base de dados dos sons o corte de cada segmento é crucial para que se possa obter uma melhor qualidade, tanto a nível segmental como supra-segmental isto porque nenhum processamento prosódico será utilizado. O software utilizado para editar e realizar o corte dos sons de fala foi o Praat [22].

Na fig. 5, temos como exemplo, o corte no início do segmento do dígito “trinta”, este deve ser sempre iniciado no valor zero de amplitude no início de um ciclo positivo. E o corte no fim do segmento do dígito, representado na fig.6, deve ser feito também quando o valor da amplitude é zero, mas no fim de um ciclo negativo. Na fig.7 está representado o segmento do dígito “trinta” completo, cortado exatamente como foi explicado anteriormente.

Quando o corte é bem efetuado, ao haver ligação dos vários segmentos de dígitos para a reprodução de qualquer número, não irão ocorrer problemas de fase e os valores de amplitude do sinal vão ligar-se em zero sem variações bruscas da amplitude.

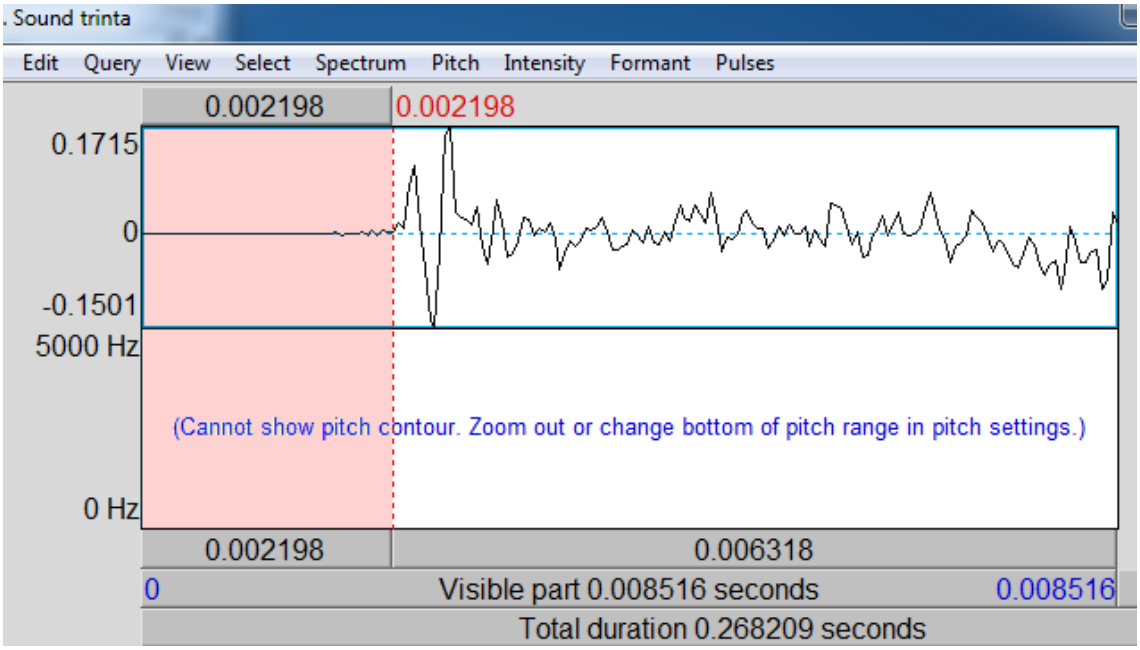

Fig. 5 – Início do corte do segmento do dígito "trinta"

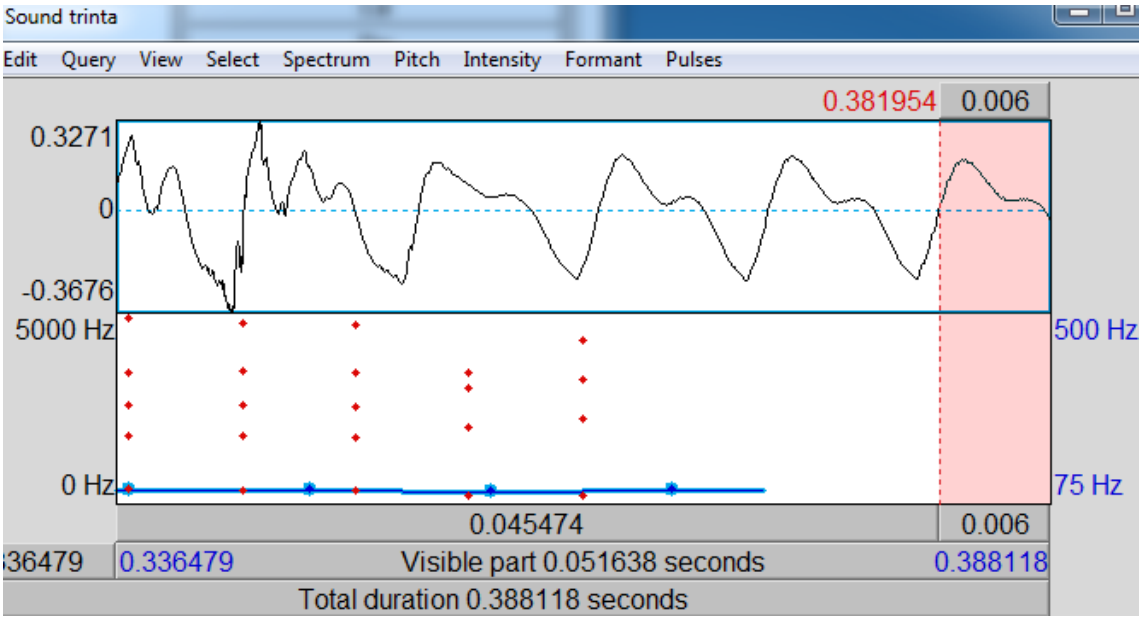

Fig.6 – Fim do corte do segmento do dígito "trinta"

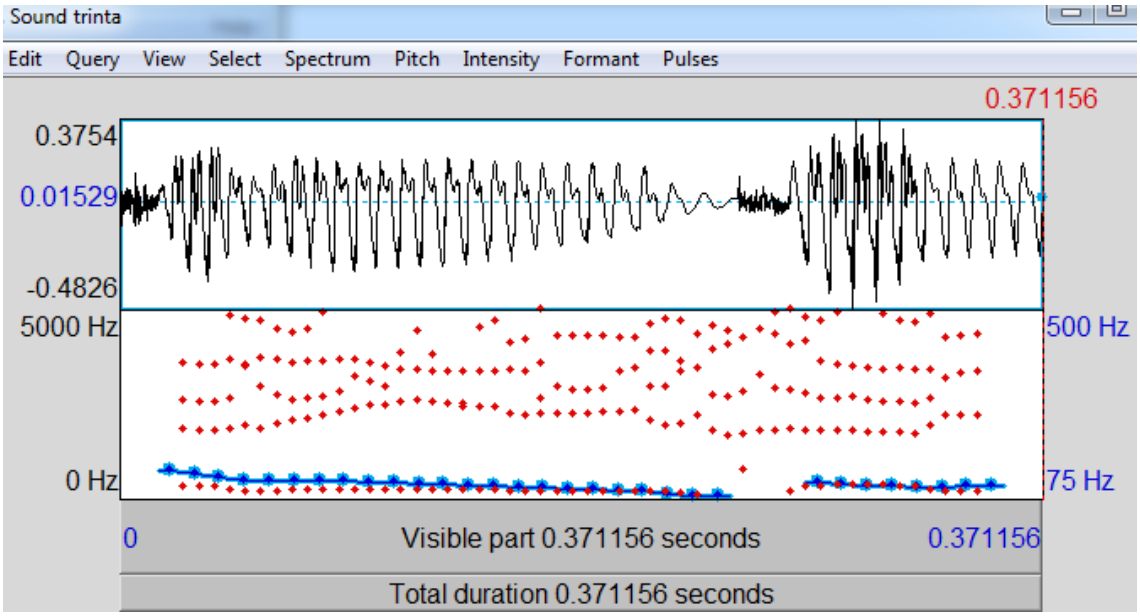

Fig.7 – Corte completo do segmento do dígito "trinta"

A prosódia de um dígito é significativamente diferente na duração e curva de F0 (entoação), dependendo da posição do dígito do número, por conseguinte, são necessárias diferentes entoações, dependendo das posições dos dígitos. Consequentemente, foram gravados e guardados registos diferentes para o mesmo dígito de um número inteiro. Um será usado numa posição inicial, outros em posições intermedias e posição final.

O som dos dígitos usados numa posição inicial ou intermédia têm uma duração menor do que os sons que estão numa posição final do número.

Temos como exemplo, o corte do segmento de voz "duzentos", numa posição inicial, fig.8. Numa posição intermédia, fig.9 o corte do segmento de voz "duzentose" (com a parte inicial do e", a seguir ao "duzentos"), que depois se irá juntar a outro número gravado com a parte final do "e" no início (ficando assim o "e" completo) e o corte do segmento de voz "duzentosf", numa posição final fig10.

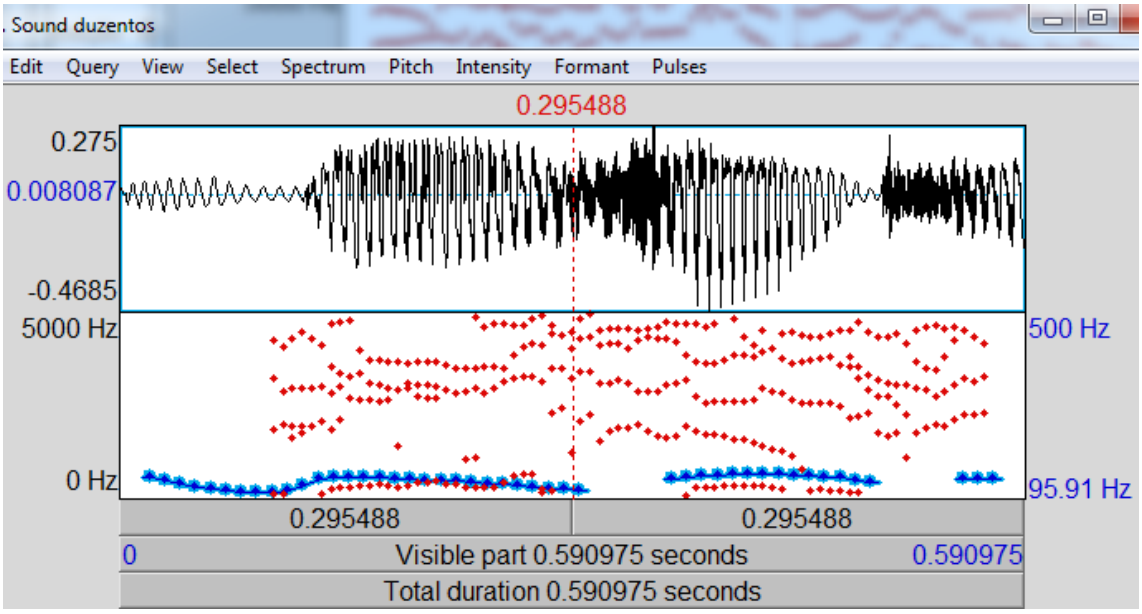

Fig.8 – Corte do segmento do dígito "duzentos" (posição inicial)

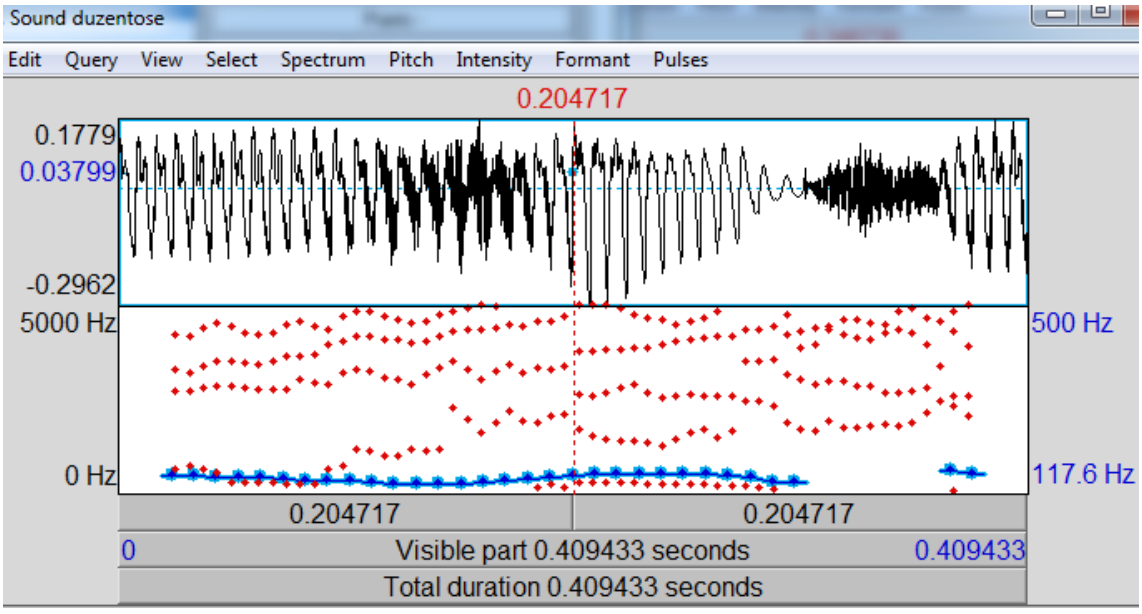

Fig. 9 – Corte do segmento do dígito "duzentose"

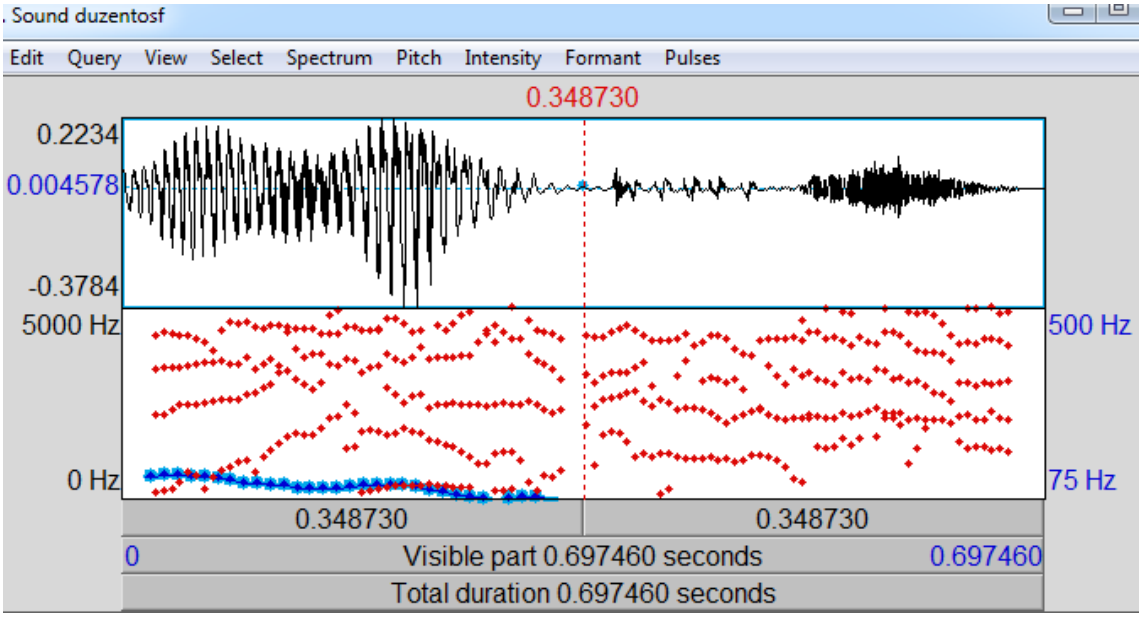

Fig. 10 – Corte do segmento do dígito "duzentosf" (posição final)

Nas figuras apresentadas anteriormente, o gráfico de cima apresenta a forma de onda acústica e a janela inferior mostra os quatro formantes. Os pontos vermelhos representam a frequência das formantes, e a linha azul representa a frequência fundamental (F0) estimada.

O segmento de uma posição final é naturalmente um som mais longo (por exemplo o da fig.10 tem 697 ms de comprimento enquanto que o da fig.8 tem apenas 591 ms de comprimento). Também o F0 denota uma diminuição no final do segmento que proporciona a informação sobre a prosódia para o ouvinte que este é o último dígito do número.

Como já foi dito anteriormente, a reprodução da leitura de um número deve ser o mais natural possível e para conseguir tal objetivo ao longo do trabalho fizeram-se algumas modificações com o objetivo de conseguir a maior naturalidade possível na conversão de um texto em fala.

Por exemplo, na leitura do número cento e sessenta e cinco "165", inicialmente foram gravados na base de dados as palavras: "cento", "sessenta", "cinco" e a partícula "e". O número era lido como cento e sessenta e cinco, deste modo ao ser reproduzido a partícula "e" ficava muito saliente, sobressaia muito em relação ao resto dos números, e

não era isso que pretendíamos. Então optamos por fazer algumas modificações no programa e na base de dados e para reproduzir o mesmo número na base de dados gravamos as palavras: "centoe", "esessentae" e "ecincof" (cada um dos "e" foi cortado a meio). Como se pode ver nas fig.11, fig.13, fig.14 e fig. 16, em que podemos ver o zoom do corte detalhado no início e no fim de cada segmento, conforme as indicações já dadas anteriormente, para não ocorrerem problemas de fase. Na fig. 12, temos a representação dos segmentos de voz "centoe", na fig. 15 a representação do segmento "esessentae", na fig. 17 está representado o segmento "ecincof" e na fig. 18 está representado o segmento da reprodução completa do número "cento e sessenta e cinco".

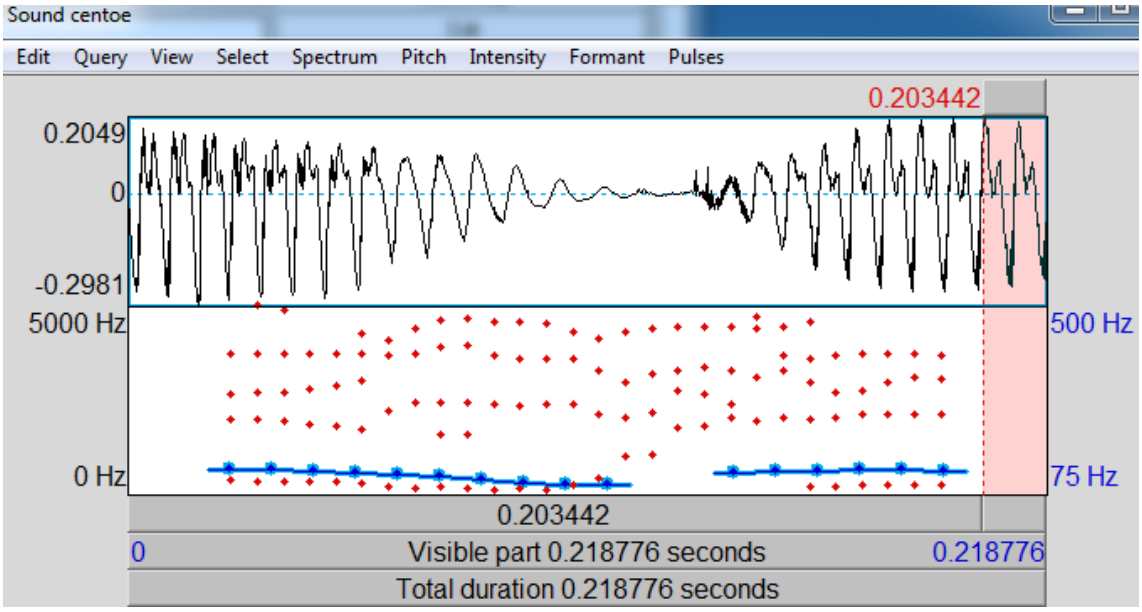

Fig.11 – Zoom do corte da parte final do segmento de voz "centoe"

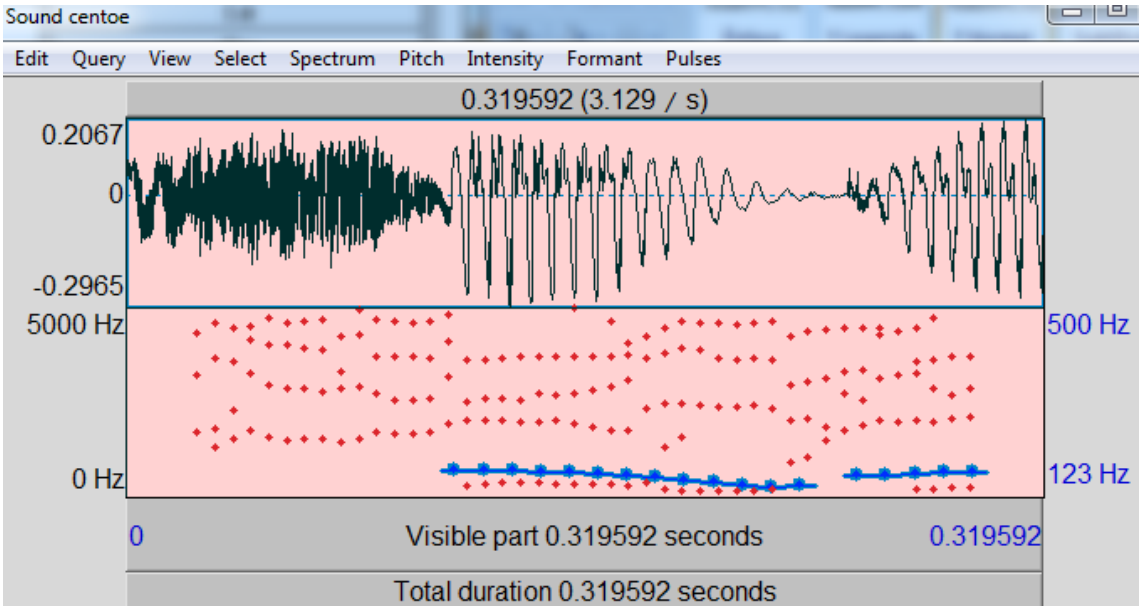

Fig.12 - Corte do segmento de voz "centoe"

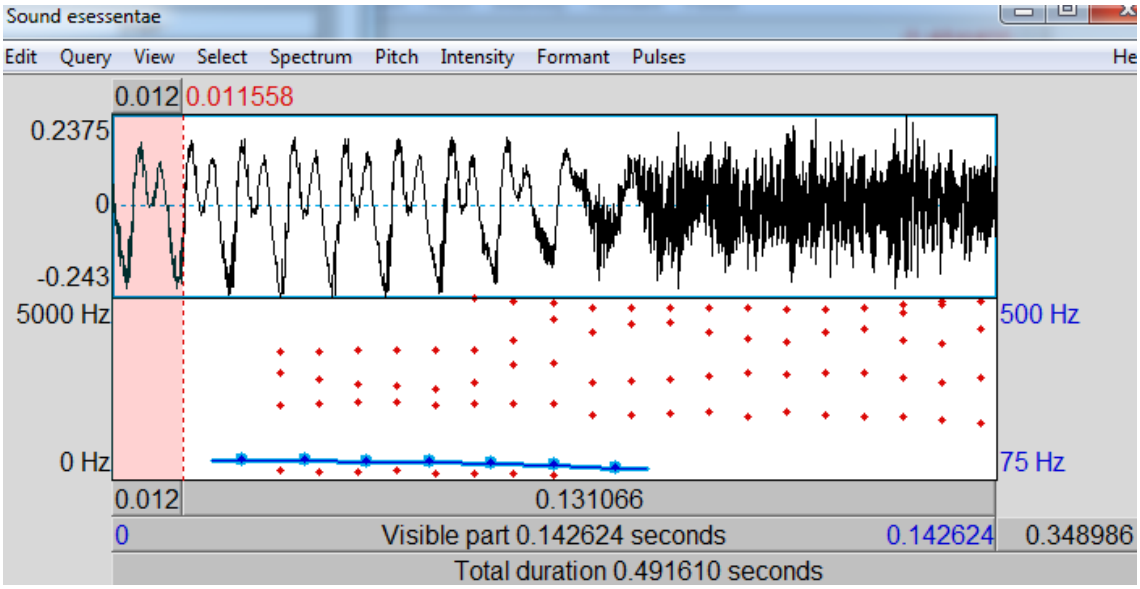

Fig.13 – Zoom do corte da parte inicial do segmento de voz "esessentae"

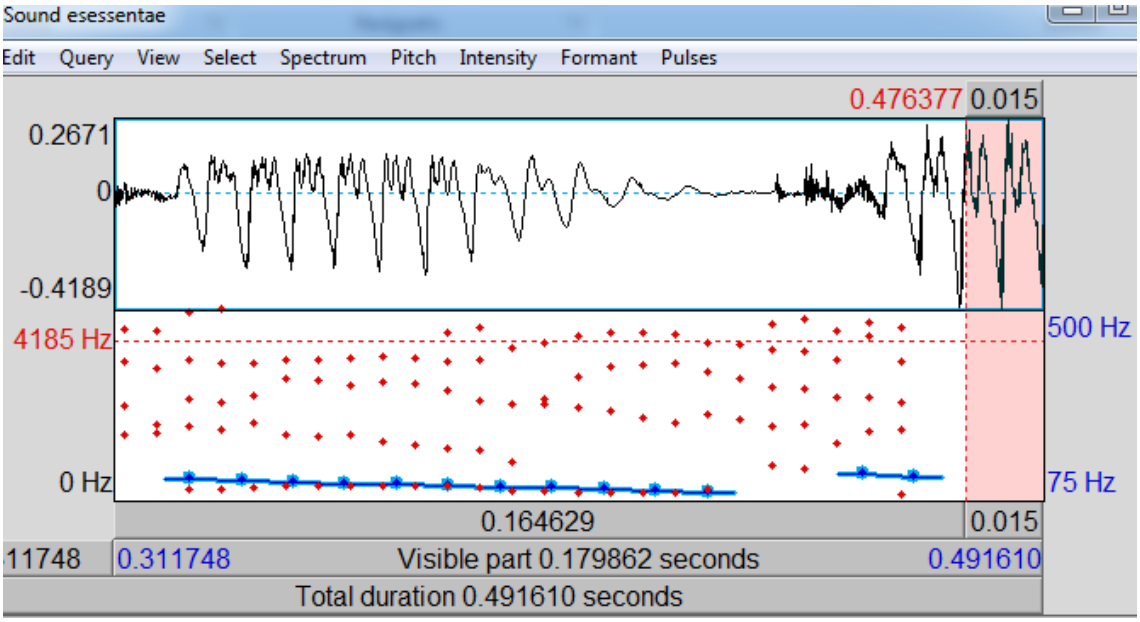

Fig.14 – Zoom do corte da parte final do segmento de voz "esessentae"

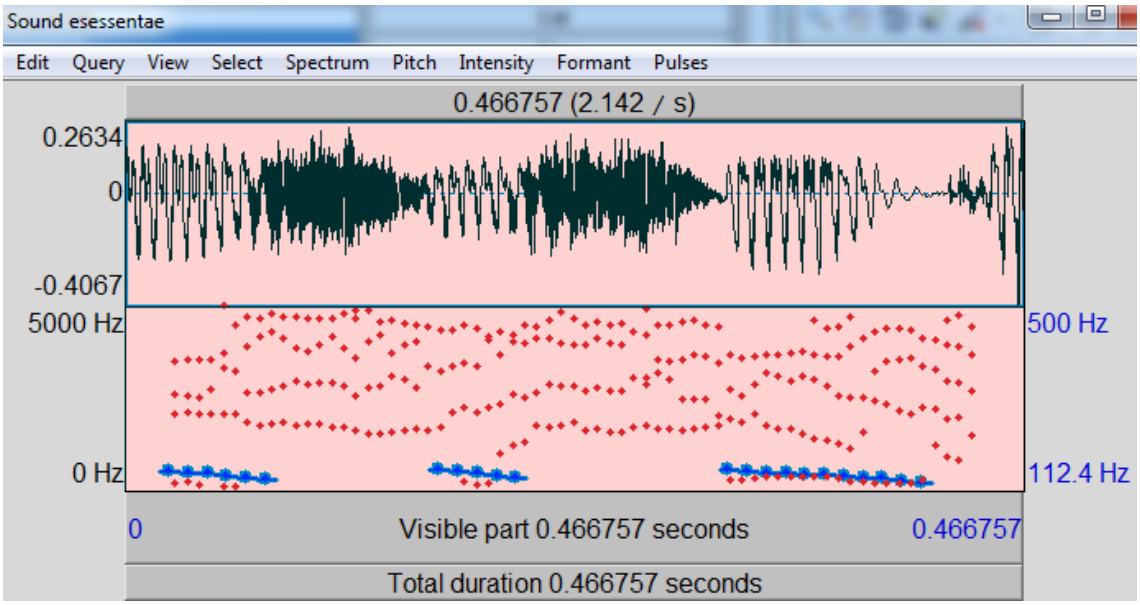

Fig.15 - Corte do segmento de voz "esessentae"

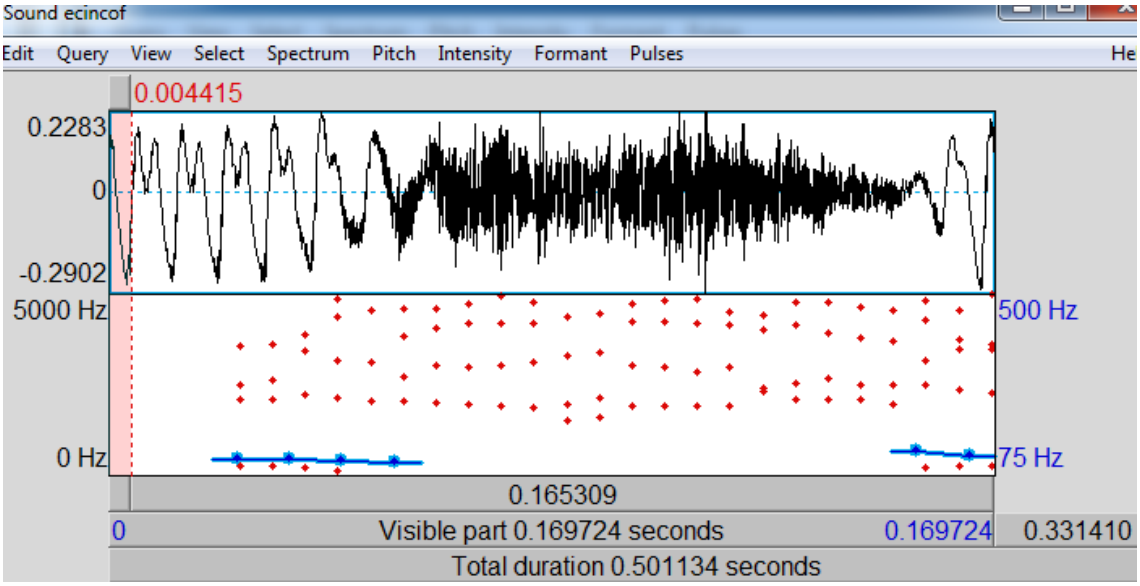

Fig.16 – Zoom do corte da parte inicial do segmento de voz "ecincof"

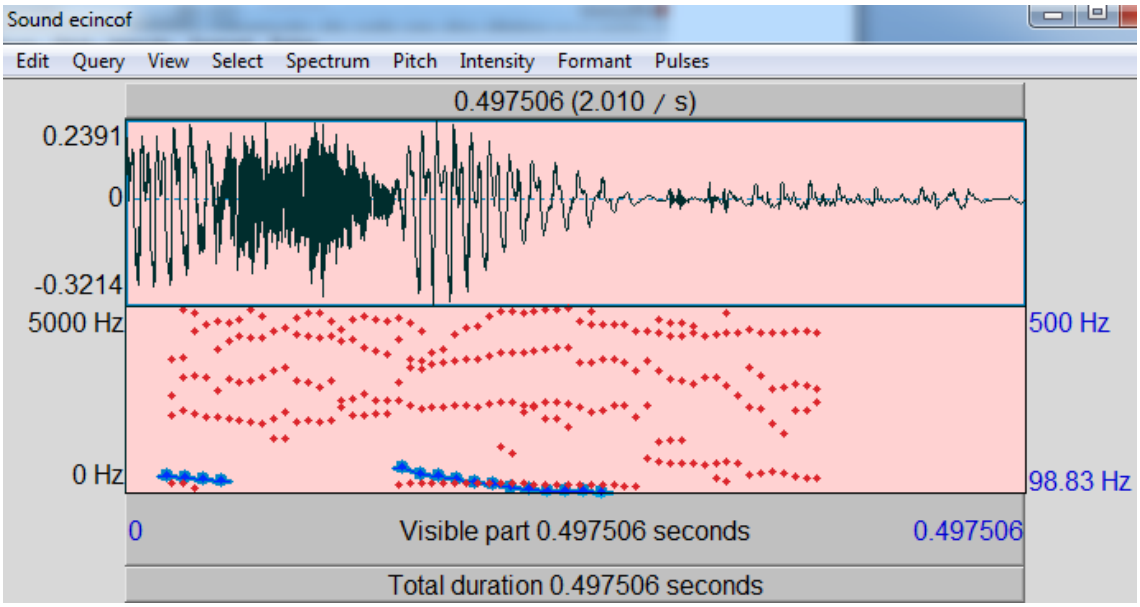

Fig.17 – Corte do segmento de voz "ecincof"

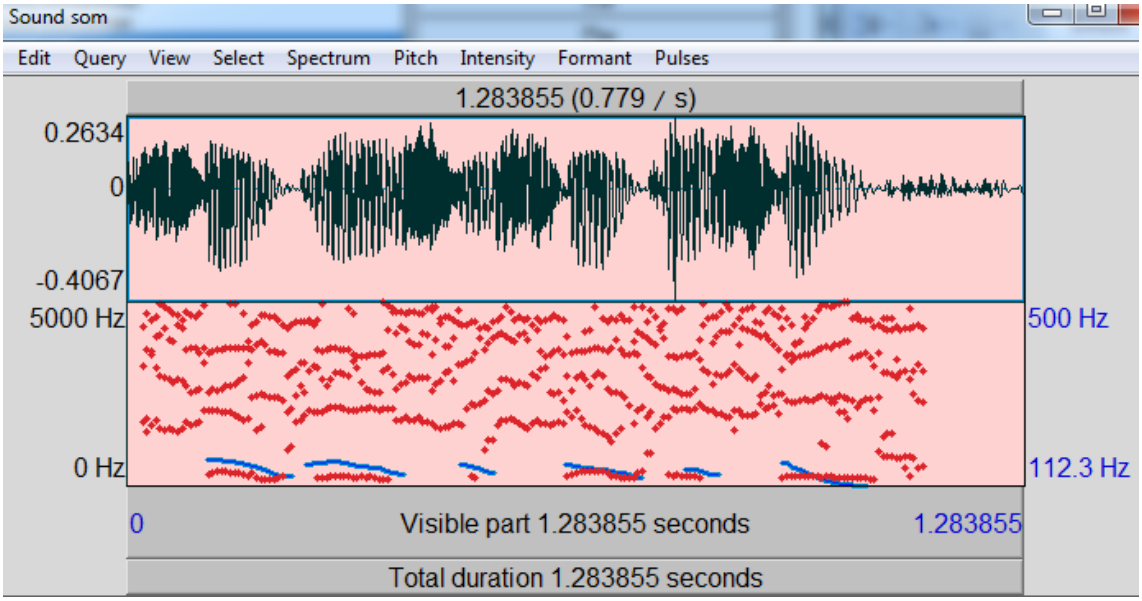

Fig.18 - Corte do som total na reprodução do número "cento e sessenta e cinco"

Como se pode verificar na fig.18, o valor da frequência fundamental ao longo da reprodução do número apresenta uma pequena variação, havendo uma diminuição do seu valor na parte final, o que nos leva a interpretar que o dígito está em posição final.

4 – Algoritmos utilizados na realização do trabalho

Foi realizado o programa principal, onde dependendo das condições vão ser chamados os programas, para ler uma data (dd-mm-aaaa), um número de telefone da rede fixa, um número de telemóvel (iniciado por 91, 92, 93 e 96), um número de identificação da segurança social, ou qualquer número entre 0 (zero) e 999 999 999 (novecentos e noventa e nove milhões, novecentos e noventa e nove mil, novecentos e noventa e nove). Na fig.19, está representado o algoritmo do programa principal. Utilizam-se algoritmos diferentes para a leitura de números de telefone fixo e números de telemóvel porque o agrupamento de dígitos que é realizado é diferente. No caso dos telemóveis faz-se um grupo com os dois dígitos iniciais e no telefone fixo faz-se um grupo com os 3/2 dígitos iniciais correspondentes ao indicativo da localidade. Como consequência, os restantes dígitos também são agrupados de forma diferente.

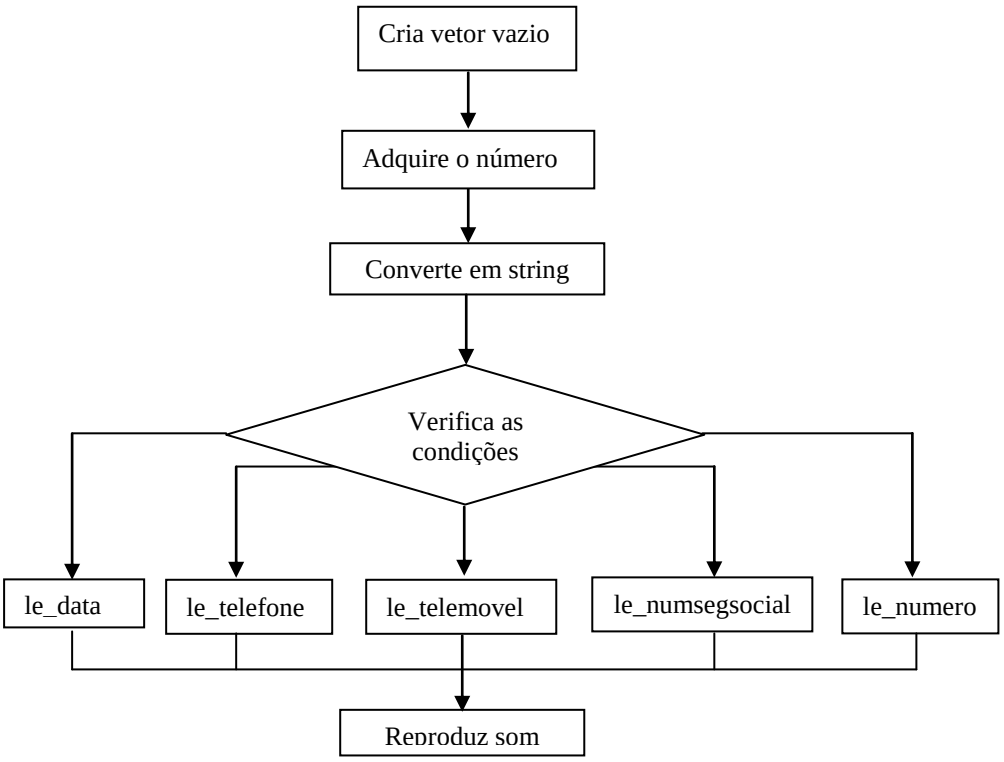

Fig.19 – Fluxograma do algoritmo do programa principal

A função principal é iniciada através da criação de um vetor vazio que recebe os segmentos a serem concatenados. Em seguida, verifica as condições do número introduzido pelo utilizador e converte numa string.

Caso sejam detetadas duas barras ou dois traços, nas posições três e seis, vai identificar o número como sendo uma data e chama a função le_data. Se o número tem nove dígitos e começa com o dígito "dois", o programa identifica o número como sendo um número de telefone, e chama a função le_telefone. Se o número tem nove dígitos e começa com os dígitos "91", "92", "93" ou "96", o programa identifica o número como sendo um número de telemóvel, e chama a função le_telemóvel. Se o número tem onze dígitos o programa identifica o número como sendo um número de identificação da segurança social e chama a função le_numsegsocial. Caso não sejam verificadas nenhumas das condições anteriores o programa vai chamar a função le_numero.

Para que o sistema fosse perfeito, deviam ser incluídas mais condições para que um número que seja uma quantidade mas que tenha a dimensão e início de um número de telefone/telemóvel ou de um n.º de segurança social pudesse ser corretamente interpretado. Futuramente este tipo de erro pode ser corrigido por outro processo como seja uma confirmação do utilizador.

4.1 - Algoritmo para os números inteiros

Este algoritmo foi desenvolvido para reproduzir números inteiros, desde o número 0 (zero) até 999 999 999 (novecentos e noventa e nove milhões, novecentos e noventa e nove mil, novecentos e noventa e nove).

Na leitura de um número com vários algarismos, fazem-se grupos de três algarismos, da direita para a esquerda. E cada grupo de algarismos representa uma classe.

A primeira classe é a das unidades a segunda classe é a dos milhares e a terceira classe é a dos milhões, como representado na figura seguinte.

cent. de milhões	dez. de milhões	unid. de milhões	cent. de milhar	dez. de milhar	unid. de milhar	Centenas	Dezenas	Unidades
5	3	2	6	9	3	4	1	7
Classe dos milhões			Classe dos milhares			Classe das unidades		

Fig.20 – Classificação dos números

Inicialmente foram gravados todos os sons dos dígitos, na base de dados, em ficheiros do tipo wav, em todas as posições para que na reprodução do número a entoação reproduzida seja a que se utiliza na linguagem naturalmente falada.

O programa principal, ou a função principal vai chamar outras funções, de acordo com as condições que se verificam a quando da introdução do dígito pelo utilizador. Os sons serão associados à respetiva variável, e carregados apenas quando são chamados. Obtém o número e cria um vetor vazio que durante o algoritmo será preenchido com a sequência de números que serão reproduzidos. Esse número será convertido para string. Com uma sequência de caracteres, é mais fácil determinar o comprimento deste número. A partir daqui este algoritmo irá associar os dígitos às variáveis de acordo com a sua posição e comprimento do número. Na fig.21 esta representado o fluxograma do algoritmo que reproduz um número.

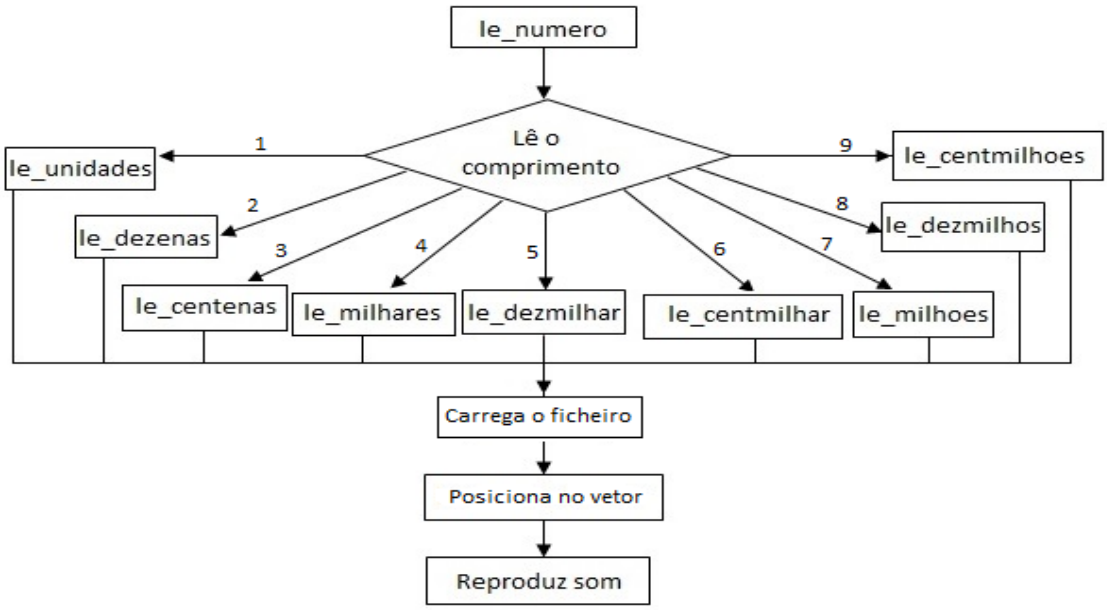

Fig.21 – Fluxograma do algoritmo da reprodução de um número

Como se pode ver na tabela IX, dependendo do tamanho do número é chamada a função correspondente, para a leitura de um número inteiro. Em seguida, associa o som à posição de cada dígito, se é um dígito de posição inicial, de posição intermedia ou de posição final. Carrega o ficheiro do arquivo da base de dados dos sons gravados, e o vetor é preenchido com a sequência de sons para depois reproduzir o som.

Tabela IX – Lista das funções para leitura de números

Comprimento do número	**Função a ser chamada**
L = 1	le_unidades
L = 2	le_dezenas
L = 3	le_centenas
L = 4	le_milhares
L = 5	le_dezmilhares
L = 6	le_centmilhar
L = 7	le_milhoes
L = 8	le_dezmilhoes
L = 9	le_centmilhoes

4.1.1 - Algoritmo para as unidades

A função le_unidades, vai ser chamada quando o programa principal deteta que vai reproduzir um número e que o seu comprimento é L=1. Os sons gravados na base de dados necessários para a reprodução das unidades são os números: zero, um, dois, três, quatro, cinco, seis, sete, oito e nove, todos eles gravados numa posição final. Na figura seguinte está representado o fluxograma do algoritmo que vai reproduzir as unidades.

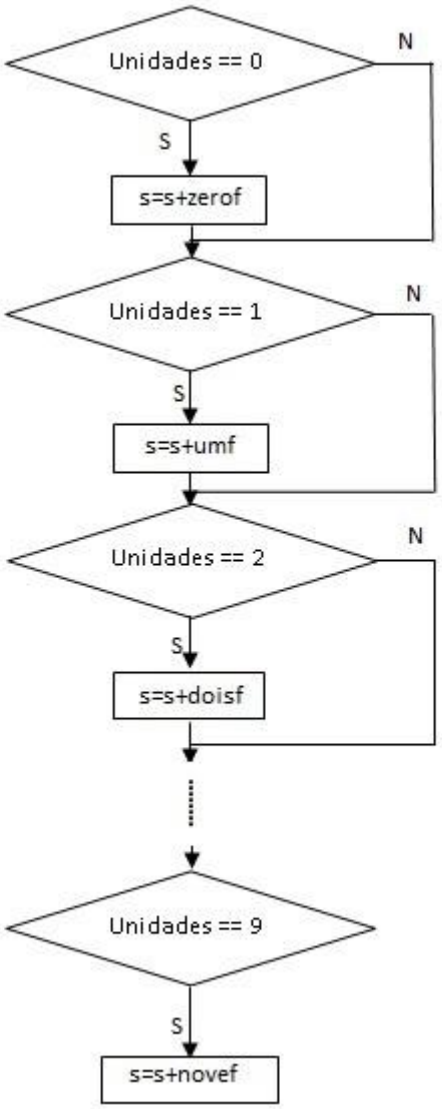

Fig.22 – Fluxograma do algoritmo para reprodução das unidades

4.1.2 - Algoritmo para as dezenas

O sistema vai reproduzir as dezenas quando o comprimento do número é L=2. Os sons gravados na base de dados necessários para a reprodução das dezenas são os números: dez, onze, doze, treze, catorze, quinze, dezasseis, dezassete, dezoito, dezanove, eum, edois, etres, equatro, ecinco, eseis, esete, eoito, enove, (com metade da partícula "e" associada no inicio do número), vinte, trinta, quarenta, cinquenta, sessenta, setenta, oitenta e noventa, estes gravados em posição final. Foram ainda gravados em posição inicial, os números vintee, trintae, quarentae, cinquentae, sessentae, setentae, oitentae, noventae, (estes com metade da partícula "e" associada, no fim do número).
Na figura seguinte está representado o fluxograma do algoritmo que vai reproduzir as dezenas.

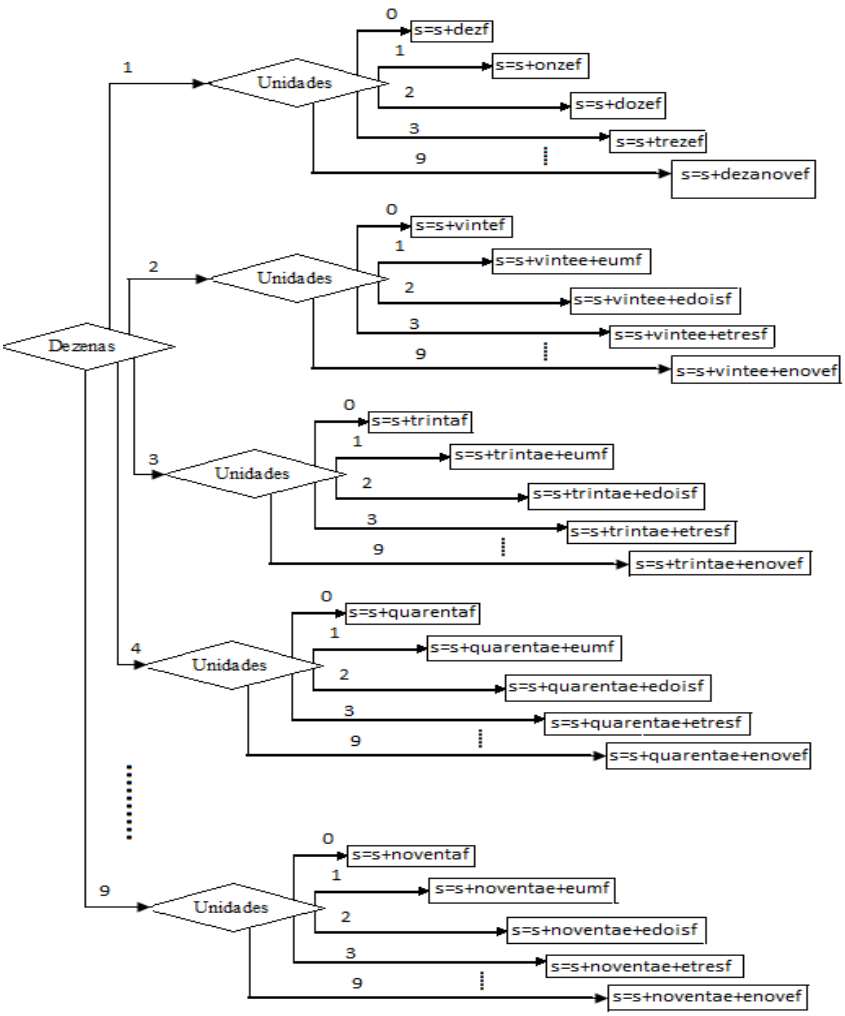

Fig.23 – Fluxograma do algoritmo para reprodução das dezenas

4.1.3 - Algoritmo para as centenas

O sistema vai reproduzir as centenas quando o comprimento do número é L=3. Os sons gravados na base de dados necessários para a reprodução das centenas são os números: edez, eonze, edoze, etreze, ecatorze, equinze, edezasseis, edezassete, edezoito, edezanove, evinte, etrinta, equarenta, ecinquenta, esessenta, esetenta, eoitenta e enoventa, (com metade da partícula "e" associada antes do número, e em posição final), cem, duzentos, trezentos, quatrocentos, quinhentos, seiscentos, setecentos, oitocentos e novecentos, estes gravados em posição final. Foram ainda gravados em posição intermédia os números: evinte, etrinta, equarenta, ecinquenta, esessenta, esetenta, eoitenta e enoventa, (com metade da partícula "e" associada antes do número), evintee, etrintae, equarentae, ecinquentae, esessentae, esetentae, eoitentae, enoventae, ecentoe, eduzentose, etrezentose, equatrocentose, equinhentose, eseiscentose, esetecentose e enovecentose, (com metade da partícula "e" associada antes e depois do número). E em posição inicial os números: centoe, duzentose, trezentose, quatrocentose, quinhentose, seiscentose, setecentose, oitocentose, e novecentose (com metade da partícula "e" associada no fim de cada número).

Na figura seguinte está representado o fluxograma do algoritmo que vai reproduzir as centenas do número.

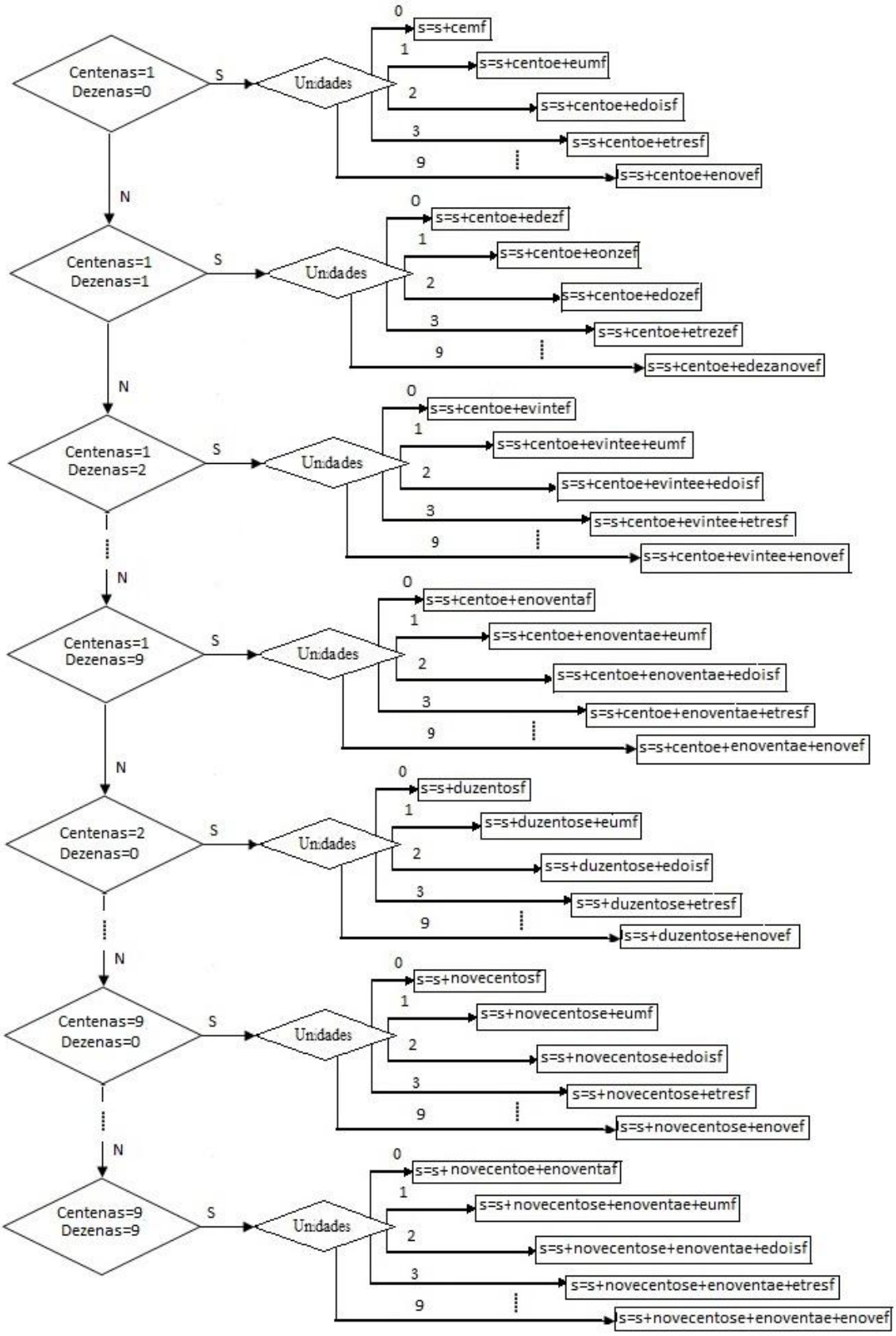

Fig.24 – Fluxograma do algoritmo para reprodução das centenas

4.1.4 – Algoritmo para os milhares e milhões

A leitura de números com comprimento superior a 3 é realizada usando os algoritmos descritos atrás para a leitura de centenas, dezenas ou unidades para ler centenas de milhares ou de milhões, dezenas de milhares ou de milhões e unidades de milhar ou de milhões. Apenas se tem que introduzir os termos "mil", "milhão" e "milhões" nas posições corretas.

Se o número é de comprimento maior do que três (L>3), esse número pode pertencer à classe dos milhares ou dos milhões. Nesse caso serão inseridos os segmentos de voz "mil", "milhão" e milhões de acordo com as seções seguintes.

4.1.4.1 Inserção de "mil" em números de comprimentos entre 4 e 6

Se o tamanho do número for igual a quatro, cinco ou seis, vai pertencer à classe dos milhares, portanto o segmento de voz "mil" vai ser reproduzido. Caso os milhares do número introduzido, seja igual a "um" reproduz "mil", se for diferente de "um", vai reproduzir esse número precedido do "mil", o mesmo acontece com as dezenas de milhar e as centenas de milhar.

4.1.4.2 Inserção de "milhão" e "milhões" em números de comprimentos entre 7 e 9

Caso o tamanho do número seja, sete, oito ou nove, vai pertencer à classe dos milhões, os segmentos de voz, "milhão" ou "milhões", serão reproduzidos. Quando a classe dos milhões for igual a "um", reproduz o segmento de voz "milhão", se for diferente de "um", vai reproduzir esse número precedido do "milhões". O mesmo acontece se o número pertencer também à classe das dezenas e centenas de milhões.
Os segmentos de voz, dos dígitos "mil", "milhares", "milhão" ou "milhões" são inseridos de acordo com a respetiva posição, o mesmo acontece com a partícula "e", como veremos mais à frente.

4.1.5 – Números em diferentes posições

Para melhorar a prosódia, cada número tem mais do que uma gravação, dependendo da posição em que se encontra. A posição vai definir o arquivo a ser lido, permitindo dar diferentes entoações dependendo da posição de cada número.

O vetor é preenchido com a sequência de sons e é reproduzido uma única vez evitando pausas entre dígitos. Verifica também se o número a carregar é ou não terminal, caso esteja na última posição.

Para números muito grandes também outras posições próximas do final exigem uma entoação particular. A gravação dos números em diferentes posições melhora este aspeto.

4.1.6 – Inserção da partícula "e"

Na maioria das funções existem condições que distinguem a leitura de alguns números, por exemplo na função le_centenas, existem condições para distinguir a leitura do numero "cem", de "cento e". Como vimos nos capítulos 4.1.2 e 41.3, metade da partícula "e" é inserida juntamente com o número, colocada atrás ou à frente do número conforme o caso. Nas funções le_milhares, le_dezmilhar, le_centmilhar, le_milhoes, le_dezmilhoes e le_centmilhoes, ocorrem situações em que a partícula "e" é adicionada sem estar associada a nenhum número especifico, por exemplo na leitura dos números 1001 (mil e um), 1100 (mil e cem), 1500 (mil e quinhentos), etc. Portanto esta secção identifica os casos em que a partícula 'e' é inserida entre as classes dos milhões, dos milhares e das unidades.

- Na função le_milhares, a partícula "e", vai ser adicionada nas situações, descritas na tabela seguinte:

Tabela X - Introdução da partícula "e" nos milhares

Situação	Exemplo	Leitura do número
X X00	2300	Dois mil **e** trezentos
X 0XX	5023	Cinco mil **e** vinte e três

- Na função le_dezmilhar, a partícula "e", vai ser adicionada nas situações descritas na tabela seguinte:

Tabela XI- Introdução da partícula "e" nas dezenas de milhares

Situação	Exemplo	Leitura do número
XX X00	21 300	Vinte e um mil **e** trezentos
XX 0XX	21 039	Vinte e um mil **e** trinta e nove

- Na função le_centmilhares, a partícula "e", vai ser adicionada nas seguintes situações:

Tabela XII- Introdução da partícula "e" nas centenas de milhares

Situação	Exemplo	Leitura do número
XXX X00	236 200	Duzentos e trinta e seis mil **e** duzentos
XXX 0XX	236 073	Duzentos e trinta e seis mil **e** setenta e três

Em resumo, sempre que há milhares e inserção do 'e' é igual independentemente dos números dos milhares, desde que aconteça uma das situações com os números centenas, dezenas e unidades dos seguintes exemplos: x00, 0xx.

- Na função le_milhoes, a partícula "e", vai ser adicionada nas seguintes situações:

Tabela XIII- Introdução da partícula "e" nos milhões

Situação	Exemplo	Leitura do número
X 000 X00	2 000 100	Dois milhões **e** cem
X 000 0XX	2 000 011	Dois milhões **e** onze
X X00 000	2 100 000	Dois milhões **e** cem mil

- Na função le_dezmilhoes, a partícula "e", vai ser adicionada nas seguintes situações:

Tabela XIV- Introdução da partícula "e" nas dezenas de milhões

Situação	Exemplo	Leitura do número
XX X00 000	23 100 000	Vinte e três milhões **e** cem mil
XX 0XX 000	23 014 000	Vinte e três milhões **e** catorze mil
XX 000 X00	23 000 300	Vinte e três milhões **e** trezentos
XX 000 0XX	23 000 025	Vinte e três milhões **e** vinte cinco

- Na função le_centmilhoes, a partícula "e", vai ser adicionada nas seguintes situações:

Tabela XV- Introdução da partícula "e" nas centenas de milhões

Situação	**Exemplo**	**Leitura do número**
XXX 000 0XX	385 000 027	Trezentos e oitenta e cinco milhões **e** vinte e sete
XXX 000 X00	385 000 100	Trezentos e oitenta e cinco milhões **e** cem
XXX 0XX 000	385 027 000	Trezentos e oitenta e cinco milhões **e** vinte e sete mil
XXX X00 000	385 100 000	Trezentos e oitenta e cinco milhões **e** cem mil

4.2 - Algoritmo para os números de telemóvel

Este algoritmo foi desenvolvido para reproduzir números de telemóvel iniciados por 91, 92, 93 e 96.

Os sons gravados na base de dados necessários para a reprodução deste tipo de números são: o número nove gravado em posição inicial (posição 9), números de zero a nove em posição intermedia (posição: 6), em posição inicial depois da pausa (posição: 2,4,7), em posição final, antes da pausa (posição 3,5,8) e em posição final (posiçãof), como mostra a fig.25. A posição define o arquivo a ser lido.

p9 p8 p7 p6 p5 p4 p3 p2 pf

— — — — — — — — —

Fig.25 – Posição de cada um dos números de telemóvel

Esta forma de agrupar os números foi a que pareceu mais adequada para os números de telemóvel que são iniciados pelos 2 primeiros dígitos que historicamente representam a rede do telemóvel. Os restantes 7 dígitos ficam melhor agrupados em grupos de 3 e de 2 dígitos de forma a evitar grupos de 4 dígitos pouco naturais. Naturalmente haveria outras formas de agrupar os dígitos, como sucede no caso dos números de telefone fixo e inclusivamente os grupos de 2 dígitos poderiam ser lidos como um número (ex. 76 - setenta e seis).

Na fig.26, está representado o fluxograma do algoritmo para reproduzir os números de telemóvel. Se o número tem nove dígitos e começa com 91, 92, 93 ou 96 o programa identifica o número como tal.

A função principal é iniciada através da criação de um vetor vazio que recebe os segmentos a serem concatenados.

Em seguida, associa o som à posição de cada dígito, se é um dígito de posição inicial, de posição intermedia, de posição inicial depois da pausa, de posição final antes da pausa ou de posição final. Carrega o ficheiro do arquivo da base de dados dos sons gravados e quando o número é reproduzido permite dar diferentes entoações dependendo da posição de cada dígito.

Os números são reproduzidos da esquerda para a direita, os dois primeiros dígitos são lidos em grupo, de seguida é lido um grupo de três dígitos e por fim dois grupos de dois dígitos, existindo uma pequena pausa entre a leitura de cada um dos grupos. O último dígito é reproduzido como sendo um dígito de uma posição final.

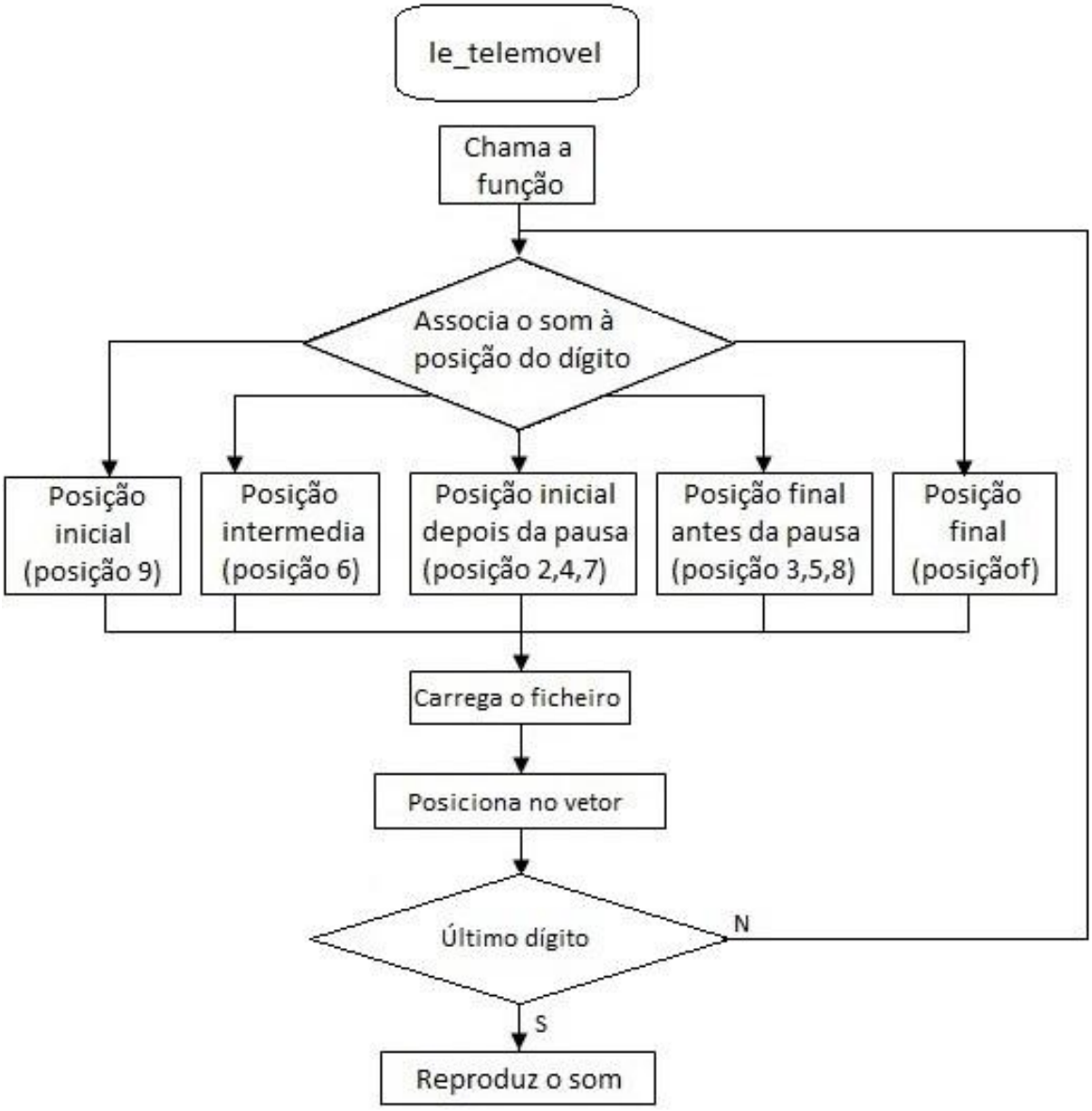

Fig.26 – Fluxograma do algoritmo que reproduz os números de telemóvel

A figura seguinte demonstra a reprodução de um número de telemóvel.

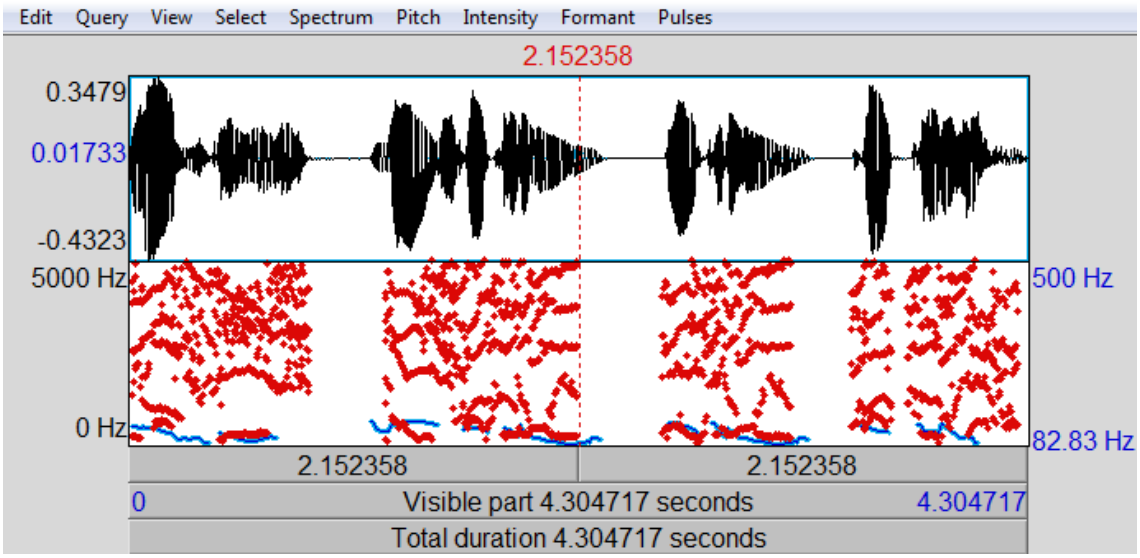

Fig.27 - Reprodução de um número de telemóvel “93 241 81 40”

4.3 - Algoritmo para os números de telefone da rede fixa

Este algoritmo foi desenvolvido para reproduzir números de telefone da rede fixa.

Os sons gravados na base de dados necessários para a reprodução deste tipo de números são: o número dois gravado em posição inicial (posição 9), números de zero a nove em posição intermedia (posição: 8,5,2), em posição inicial depois da pausa (posição: 6,3), em posição final, antes da pausa (posição 7,4) e em posição final (posiçãof), como mostra a fig.28. A posição define o arquivo a ser lido.

p9 p8 p7 p6 p5 p4 p3 p2 pf

— — — — — — — — —

Fig.28 – Posição de cada um dos números de telefone

Também neste caso a forma de agrupar os dígitos poderia ser diferente. A opção por esta forma considera que os 3 primeiros dígitos se referem ao histórico indicativo da região. Nos casos das regiões de Lisboa e Porto esse indicativo será constituído apenas por 2 dígitos e a razão apontada já não tem sentido, contudo o agrupamento de 3 dígitos mantém-se por defeito.

Se o número tem nove dígitos e começa com “dois”, o programa identifica o número como sendo de um telefone fixo.

A função principal é iniciada através da criação de um vetor vazio que recebe os segmentos a serem concatenados.

Em seguida, associa o som à posição de cada dígito, se é um dígito de posição inicial, de posição intermedia, de posição inicial depois da pausa, de posição final antes da pausa ou de posição final. Carrega o ficheiro do arquivo da base de dados dos sons gravados e quando o número é reproduzido permite dar diferentes entoações dependendo da posição de cada dígito.

Analisando cada posição é adicionado ao vetor o dígito correspondente. A forma como o número de telefone é reproduzido é simples: a partir da esquerda para a direita, o número é lido dígito a dígito, em grupos de três. Na fig.29, está representado o fluxograma do algoritmo para reproduzir os números de telefone.

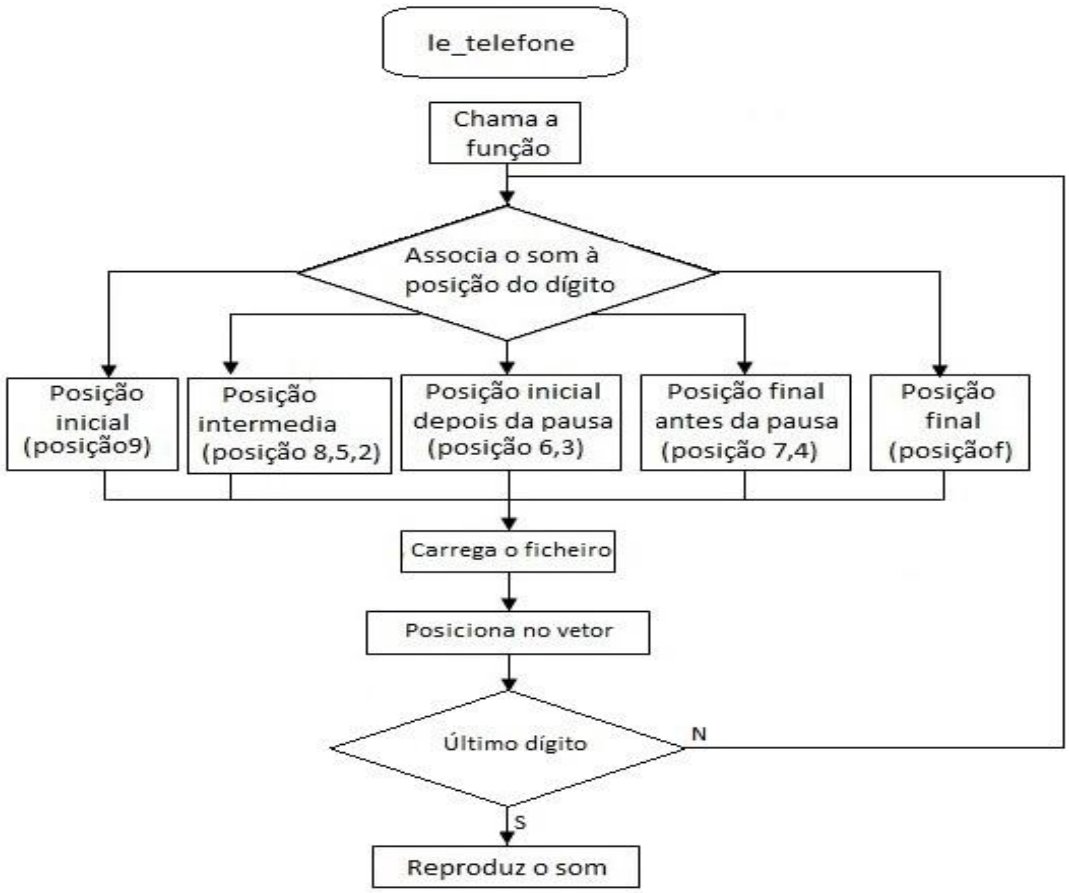

Fig.29 – Fluxograma do algoritmo que reproduz os números de telefone

A figura seguinte demonstra a reprodução de um número de telefone da rede fixa.

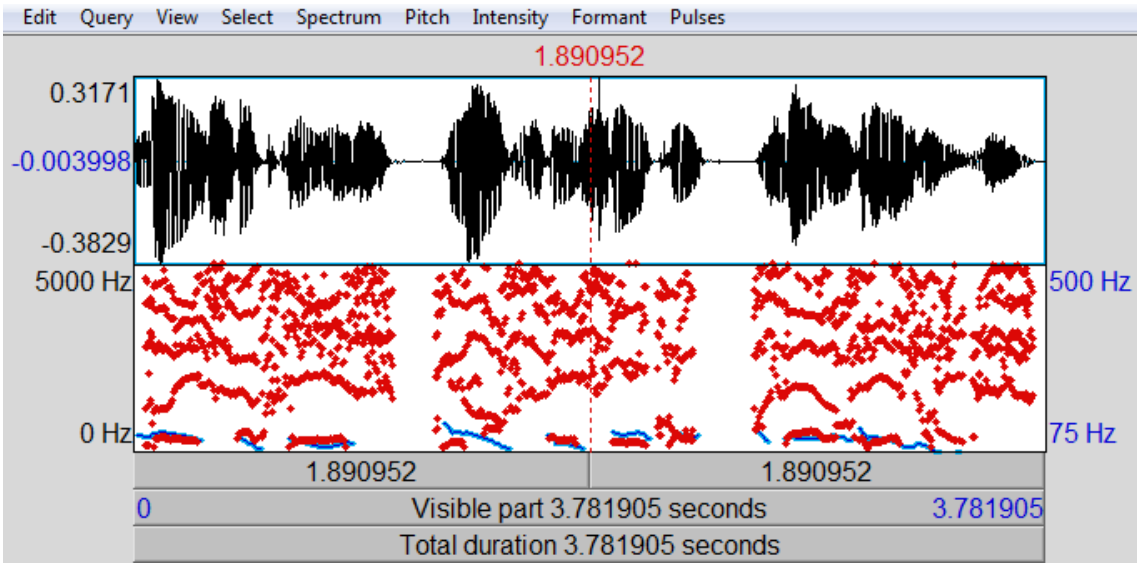

Fig. 30 – Reprodução de um número de telefone “273 965 032”

4.4 - Algoritmo para reprodução de um número da Segurança Social

Este algoritmo foi desenvolvido para reproduzir o número de identificação da segurança social, o mesmo algoritmo pode ser usado para reproduzir números de identificação fiscal, números do cartão de cidadão, números de identificação bancária, entre outros.

Os sons gravados na base de dados necessários para a reprodução deste tipo de números são: números de zero a nove, gravados em várias posições, inicial, intermedia e em posição final. Em posição inicial (posição:11), em posição intermedia (posição: 4,7,10), em posição inicial depois da pausa (posição: 2,5,8), em posição final, antes da pausa (posição 3,6,9) e em posição final (posiçãof), como mostra a fig.31. A posição define o arquivo a ser lido.

p11 p10 p9 p8 p7 p6 p5 p4 p3 p2pf

— — — — — — — — — — —

Fig.31 – Posição de cada um dos números de identificação da Segurança Social

Se o número tem onze dígitos e começa com "120" o programa identifica-o como sendo número da segurança social, correndo o risco de poder não ser. Futuramente este erro pode ser corrigido por outro processo como seja uma confirmação do utilizador.

A função principal é iniciada através da criação de um vetor vazio que recebe os segmentos a serem concatenados.

Em seguida, associa o som à posição de cada dígito, se é um dígito de posição inicial, de posição intermedia, de posição inicial depois da pausa, de posição final antes da pausa ou de posição final. Carrega o ficheiro do arquivo da base de dados dos sons gravados e quando o número é reproduzido permite dar diferentes entoações dependendo da posição de cada dígito.

O número é reproduzido dígito a dígito, da esquerda para a direita, em três grupos, de três dígitos e um último grupo de dois dígitos. O último dígito reproduzido é lido como sendo um dígito em posição final.

Na fig.32, está representado o fluxograma do algoritmo para reproduzir os números de identificação da segurança social.

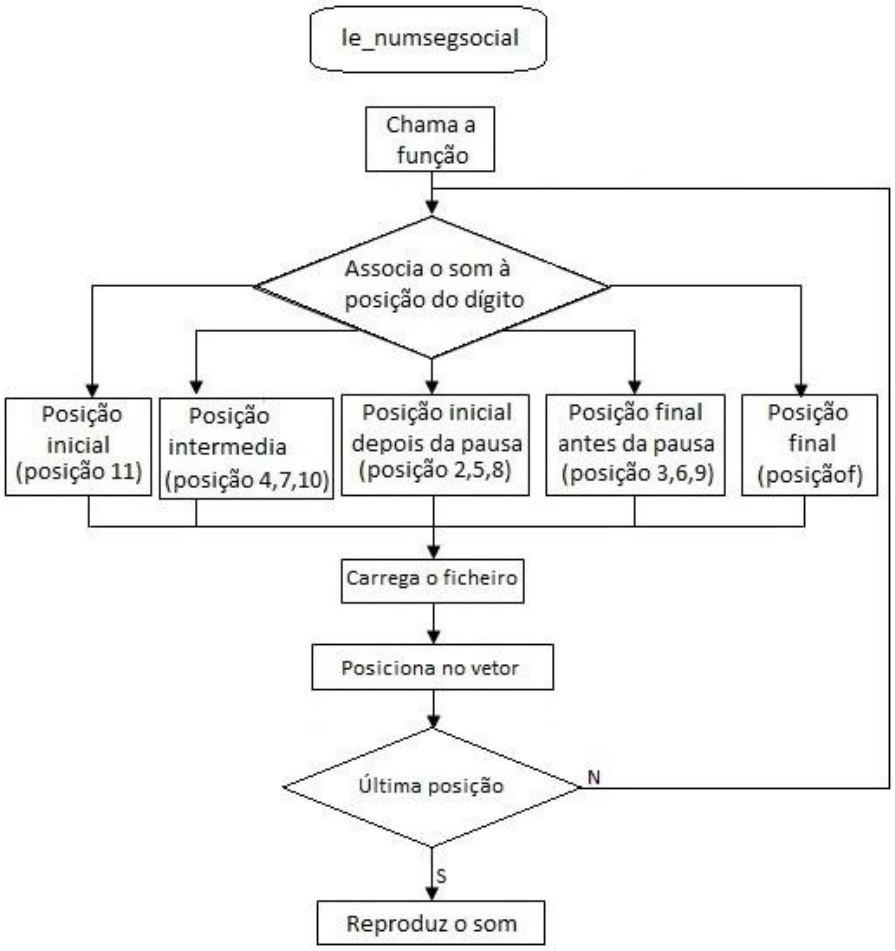

Fig.32 – Fluxograma do algoritmo que reproduz os números da Segurança Social

A figura seguinte demonstra a reprodução de um número de identificação da Segurança Social.

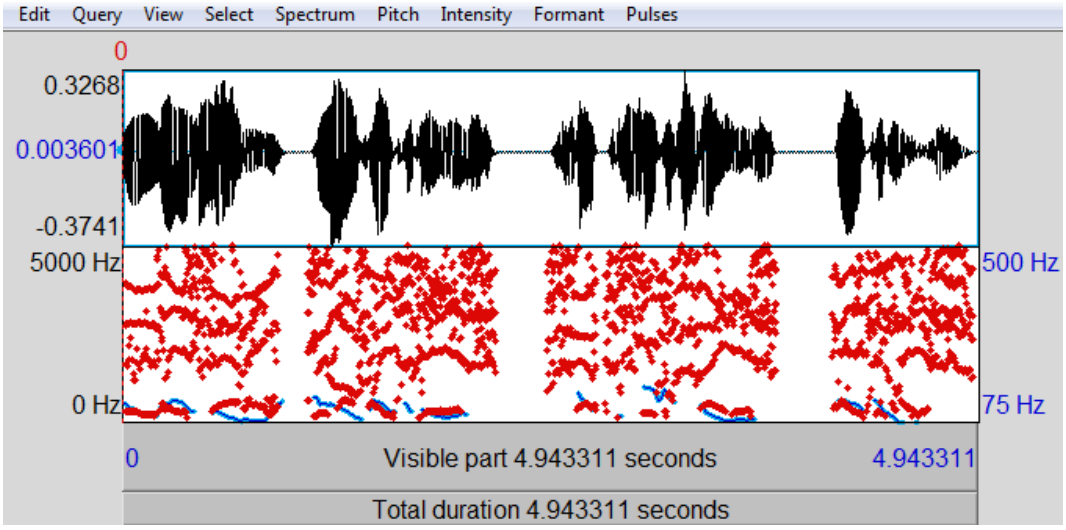

Fig. 33 – Reprodução de um número de identificação da Seg. Social "120 893 756 82"

4.5 - Algoritmo para reprodução das datas (dd-mm-aaaa)

Este algoritmo foi desenvolvido para reproduzir datas no formato (dd-mm-aaaa) ou (dd/mm/aaaa).

Para uma data, além dos sons já mencionados que foram gravados na base de dados, foram também gravadas as partículas "do" e "de". Por exemplo esta data: 06-04-2014, vai ser lida como "seis, do quatro, de dois mil e catorze".

A função principal é iniciada através da criação de um vetor vazio que recebe os segmentos a serem concatenados.

Em seguida, associa o som à posição de cada dígito. Carrega o ficheiro do arquivo da base de dados dos sons gravados e quando a data é reproduzida permite dar diferentes entoações dependendo da posição de cada dígito.

A data vai ser lida da esquerda para a direita, começado pelo dia, mês e ano. Na fig.34, está representado o fluxograma do algoritmo para reproduzir as datas.

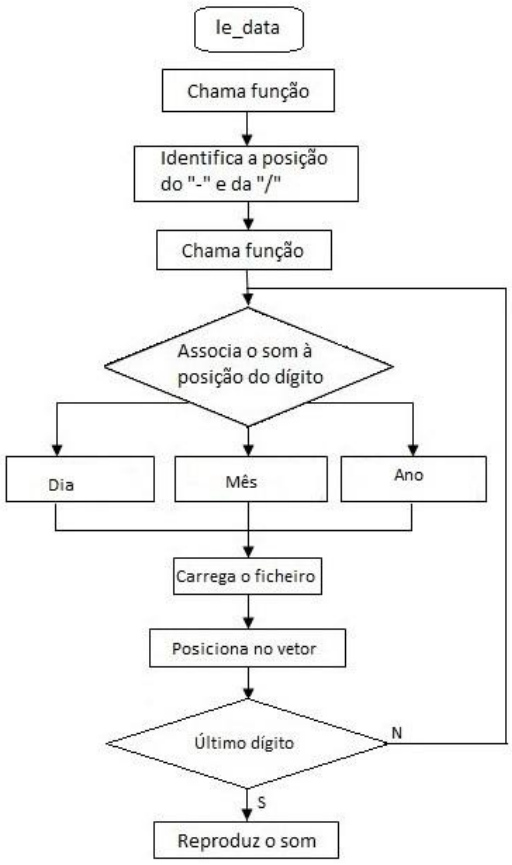

Fig.34 – Fluxograma do algoritmo que reproduz datas

A figura seguinte demonstra a reprodução de uma data.

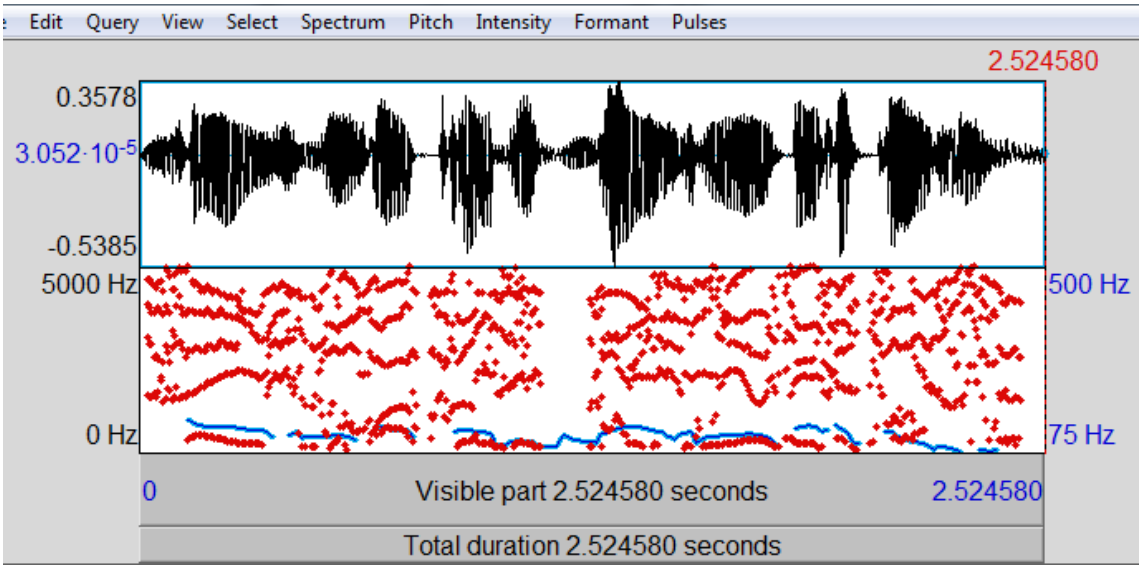

Fig. 35 – Reprodução da data “seis, do quatro, de dois mil e catorze”

5 – Análise dos resultados

Para avaliar a qualidade do trabalho realizado foi efetuado um teste percetual, para medir a qualidade dos algoritmos. Para isso foi usada uma escala de 1 a 5. Em que 1 – Muito Mau; 2 – Mau; 3 – Razoável; 4 – Bom e 5 - Excelente.

Foram submetidos 15 indivíduos à realização do teste, ouviram a reprodução de vinte números, em quatro versões diferentes do trabalho. A versão 1, da fase inicial do trabalho, a versão 2, noutra fase intermédia, a versão final do trabalho e por fim um conjunto de números lidos diretamente pelo locutor. Esta última versão lida pelo locutor permitia que fosse possível aos ouvintes terem um termo de comparação e assim avaliar com uma referência a reprodução de cada número. De salientar que a 1ª e 2ª versão, tinha uma base de dados gravada com uma voz feminina, não profissional e na versão final como já foi referido anteriormente a base de dados foi gravada com um profissional. Cada versão englobava os cinco algoritmos e cada número foi reproduzido uma única vez, numa sequência aleatória, nas quatro versões diferentes, num total de oitenta reproduções. Em seguida, os indivíduos foram convidados a avaliar cada reprodução considerando a qualidade do som, a perceção auditiva e a exatidão das palavras usadas (números e partículas) para produzir o número.

Os números utilizados no teste são apresentados na Tabela XVI, a segunda coluna representa a forma correta de cada reprodução e a terceira coluna o algoritmo em que se insere.

Após análise estatística do teste de opinião os resultados são apresentados na Tabela XVII, onde se verifica uma evolução no trabalho, da versão1 para a versão2 e depois para a versão final, onde os resultados são bastante positivos, o Mean Opinion Score (MOS) de 4,46, de 4,45 para o algoritmo dos números, 4,67 para o algoritmo dos números de telemóvel, 4,73 para o algoritmo dos números de telefone, 4,60 para o algoritmo de identificação do número da Segurança Social e 4,13 para o algoritmo das datas. De salientar que na reprodução dos números de telemóvel o sistema na versão final foi inclusivamente melhor avaliado que a voz do locutor, 4,67 e 4,50, respetivamente. Os restantes resultados da avaliação foram bastante próximos dos resultados, em que o locutor lê diretamente os números.

Verifica-se também uma evolução bastante positiva nos resultados da média de cada algoritmo, nas quatro versões e com o nível geral de muito boa qualidade. Todos os números testados foram reproduzidos corretamente, sem erros de processamento.

Tabela XVI - Números reproduzidos

Número	Reprodução do número	Algoritmo
28	Vinte e oito	Números
135	Cento e trinta e cinco	Números
2 367	Dois mil trezentos e sessenta e sete	Números
21 045	Vinte e um mil e quarenta e cinco	Números
56 897	Cinquenta e seis mil, oitocentos e noventa e sete	Números
348 603	Trezentos e quarenta e oito mil, seiscentos e três	Números
475 234	Quatrocentos e setenta e cinco mil, duzentos e trinta e quatro	Números
1 234 056	Um milhão, duzentos e trinta e quatro mil e cinquenta e seis	Números
3 052 007	Três milhões, cinquenta e dois mil e sete	Números
73 405 784	Setenta e três milhões, quatrocentos e cinco mil, setecentos e oitenta e quatro	Números
902 008 404	Novecentos e dois milhões, oito mil, quatrocentos e quatro	Números
901 000 876	Novecentos e um milhões, oitocentos e setenta e seis	Números
506 071 002	Quinhentos e seis milhões, setenta e um mil e dois	Números
987 654 321	Novecentos e oitenta e sete milhões, seiscentos e cinquenta e quatro mil, trezentos e vinte e um	Números
93 765 84 95	Nove três, sete seis cinco, oito quatro, nove cinco	Telemóvel
96 837 54 18	Nove seis, oito três sete, cinco quatro, um oito	Telemóvel
276 965 032	Dois sete seis, nove seis cinco, zero três dois	Telefone
120 983 756 82	Um dois zero, nove oito três, sete cinco seis, oito dois	NºSeg. Social
02-05-2002	Dois, do cinco, de dois mil e dois	Data
03/06/2009	Três, do seis, de dois mil e nove	Data

Tabela XVII – Média das opiniões dos sujeitos

	Algoritmo dos números	**Algoritmo dos números de telemóvel**	**Algoritmo dos números de telefone**	**Algoritmo do número da Seg.Social**	**Algoritmo das datas**	**MOS**
Versão 1	2,14	2,40	1,53	2,73	2,93	**2,26**
Versão 2	3,27	2,76	2,07	2,93	3,37	**3,14**
Versão Final	**4,45**	**4,67**	**4,73**	**4,60**	**4,13**	**4,46**
Locutor	4,76	4,50	4,86	4,67	4,67	**4,72**
Número de avaliações	210	30	15	15	30	____

6 – Conclusões e desenvolvimentos futuros

Este capítulo apresenta de forma resumida as conclusões do trabalho desenvolvido. Pretende-se também deixar um alerta para as linhas de investigação que ficam em aberto neste trabalho e fazer um apontamento de algumas aplicações ligadas ao tema desenvolvido.

6.1- Conclusões

O objetivo deste trabalho foi desenvolver um sistema de leitura automática que permita a leitura de números entre 0 (zero) e 999 999 999 (novecentos e noventa e nove milhões, novecentos e noventa e nove mil, novecentos e noventa e nove), números de telemóvel, números de telefone, número de identificação da segurança social e datas no formato (dd-mm-aaaa).

Foram desenvolvidos cinco algoritmos diferentes, cada um com sua própria estrutura, funções e características.

Apesar das pequenas diferenças existentes entre eles, os resultados de um teste de perceção dão uma média no nível muito bom. O algoritmo para reproduzir números de telefones foi o que teve a opinião média mais elevada, **4,73**. A opinião média total foi de **4,46** num máximo de 5.

Para os sistemas de síntese é importante começar com boa qualidade dos sons gravados e com cortes precisos, caso contrário deixará de fornecer bons resultados. O locutor deve ter uma voz clara e uma agradável dicção. Por essa razão os sons foram gravados por um locutor profissional, onde foram gravados números em diferentes posições, com e sem a partícula "e" associada no início ou no fim do número.

Os algoritmos precisam de levar em consideração vários números específicos para inserir todas as condições necessárias para uma reprodução perfeita. Não menos importante é que a reprodução ocorre apenas uma vez para evitar pausas entre os dígitos, unidades e partículas.

A síntese é baseada na concatenação de segmentos de voz gravada.

Geralmente, a síntese por concatenação de segmentos longos produz um som mais natural que nos casos de voz sintetizada, porque nenhumas mudanças são necessárias nos segmentos de sons. Isto é um dos princípios do método de síntese por seleção de unidades. No entanto, devido à possibilidade de ocorrer diferenças de fase ou magnitude da onda acústica dos sons concatenados, pode, por vezes, resultar em falhas audíveis na saída.

Neste trabalho, é utilizada a síntese específica no domínio do tempo. O sistema faz a concatenação de palavras e números pré-gravados. É bastante utilizado em aplicações onde a variedade de testes a ser produzido é limitado a um determinado domínio conhecido previamente, como anúncios de agendamento de trânsito ou até mesmo previsões de tempo. Tem uma utilização comercial por um longo período de tempo, em dispositivos de voz tais como, relógios e calculadoras. O nível de naturalidade destes sistemas pode ser muito elevado, porque existe uma variedade de tipo de frases limitada, e coincidem com a prosódia e entoação das gravações originais.

Estes sistemas são limitados pelo número de palavras e frases existentes na base de dados, uma vez que não são de uso geral e só podem sintetizar as combinações de palavras e números com os quais tenham sido pré programados.

Durante a realização do trabalho foram muito importantes os cuidados tidos na gravação de números em diferentes posições (inicial, intermédia e final) para serem usados nessas mesmas posições. Revelou-se também de grande importância para a melhoria da qualidade final a gravação de números com metade da partícula “e” antes do número, depois do número e antes e depois do número, para serem usados quando é necessário incluir a partícula “e” de forma a minimizar os problemas inerentes ao corte dos sons. Teve-se ainda especial cuido na gravação da base de dados para que o locutor profissional impusesse pausas em palavras em posição final obrigando a uma paragem de cerca de 10 segundos entre cada gravação.

De salientar ainda que o corte dos segmentos de som foi realizado coerentemente de forma a evitar problemas de fase, como já foi explicado no capítulo 3. Também se teve um cuidado particular na gravação da base de dados para que o nível de entoação (F0) fosse muito próximo em segmentos concatenados e o nível de amplitude também é semelhante de forma a não ocorrerem grandes diferenças de amplitude quando se concatenam segmentos de origens diferentes.

6.2 – Desenvolvimentos futuros

Pode-se melhorar o sistema de voz sintetizada, introduzindo algumas modificações prosódias. Isso permite que os ouvintes, por exemplo, reconheçam qual é a posição do dígito em cada número.

Utilizar algoritmos de modificação prosódica que podem alterar o valor da frequência fundamental (F0), de segmentos com tons diferentes e tornar esses tons próximos, quando os segmentos de voz ficam juntos. Isto poderá melhorar muito a locução do número em que os segmentos têm tons com níveis muito diferentes. Neste trabalho esse problema foi minimizado escolhendo segmentos com níveis de F0 muito próximos e também foram muito importantes os cuidados já referidos anteriormente que se tiveram na gravação da base de dados.

Bibliografia

[1] – Ferreira, Helder Filipe "Leitura Automática De Expressões Matemáticas Audiomath", dissertação de Mestrado em Engenharia Informática na Faculdade de Engenharia da Universidade do Porto, 2005.

[2] – Cal Coimbra (2005), Artigo - "Análise Acústica da Voz".
http://www.acessa.com/viver/arquivo/vida_saudavel/2005/01/14-cal/

[3] – Machado, A. 2003 "Neuroanatomia Funciona" Editora Atheneu

[4] – Teixeira, João P.; Joaquim Silva; José Dias, Pedro Conceição and Pedro Freitas, Automatic System of Reading Numbers. International Journal on Signal & Image Processing, Vol. 5, pp 13-20, May 2014.

[5] – Teixeira, João Paulo "Modelização Paramétrica de Sinais Para Aplicação em Sistemas de Conversão Texto-Fala", dissertação de Mestrado em Engenharia Eletrotécnica e de Computadores na Faculdade de Engenharia da Universidade do Porto, 1995.

[6] – "Fonética, Fonologia e Morfologia do Português", M. H. Mira Mateus, A. Andrade, M. do Céu Viana, A. Villalva, Universidade Aberta Lisboa 1990.

[7] – "Ouvir Falar – Introdução à Fonética da Português", M. R. Delgado Martins, segunda edição, Caminho Colecção Universitária série Linguística 1992.

[8] – "Estudos de Fonologia Portuguesa", Regina C. P. Silveira, Cortez Editora 1986.

[9] – Campbell, W.N., Timing in Speech: A Multi-Level Process, In "Prosody: Theory and Experiment". Edited by Merle Horne, Kluwer Academic Publishers, pages 281-334, 2000.

[10] – Barbosa P., Bailly G., Characterisation of rhythmic patterns for text-to-speech synthesis, in Speech Comunication, 15:127-137, 1994.

[11] – Teixeira, J.P., Prosody Generation Model or TTS Systems – Segmental Durations and F0 Contours with Fujisaki Model. LAP LAMBERT Academic Publishing ISBN-13:978-3-659-16277-0, 2012.

[12] – Fujisaki, H., Dynamic characteristics of voice fundamental frequency in speech and singing. In Mac Neilage. In P. F., Editor. The Production of Speech, pages 39-55. Springer-Verlang, 1983.

[13] – Pierrehumbert, J.B. The Phonology and Phonetics of English Intonation. PhD thesis, Massachusetts Istitute of Technology, 1980.

[14] – Taylor P., "Analysis and Synthesis of Intonation using the Tilt Model". Journal of the Acoustical Society of America. Vol 1073, pp 1697-1714, 2000.

[15] – Hirst, D. and Di Cristo, A., Intonation Systems – A Survey of Twenty Languages. Cambridge University Press, 1998.

[16] – Klatt, D.H., "Software for a cascade/parallel formant synthetizer". Journal of the Acouustical Society of America, 67:971-995, 1980.

[17] – Marques, J.S.S., Modelamento Sinusoidal da Fala – aplicação à codificação a ritmos médios e baixos. PhD thesis – Instituto Superior Técnico, Lisbon, 1990.

[18] – Charpentier, F e Moulines, E., "Pitch –synchronous waveforme processing techniques for texto-to-speech synthesis using diphones". Speech Communication, 9(5/6):452-467,1990.

[19] – Silva, C. A., Automatic Extraction of the Parameters of an Articulatory Model for Speech Syntheis, PhD Thesis, DEI-University oh Minho, Portugal, 2001.

[20] – A. Conkie, "A robust unit selection system for speech synthesis." In: Proc. 137th meet. ASA forum Acusticum, Berlin, March 1999.

[21] - T. Yoshimura, K. Tokuda, T. Masuko t. Kobayashi, and T. Kitamura, "Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis," in Proc. EuroSpeech 1999, vol.5, 1999, pp. 2347-2350.

[22] - Boersma, Paul and Weenink, David. "Praat: doing phonetics by computer". Phonetic Sciences, University of Amsterdam. http://www.fon.hum.uva.nl/praat/

Anexos

Anexo 1: Lista dos sons gravados na base de dados

Unidades

zerof; umf; doisf; tresf; quatrof; cincof; seisf; setef; oitof; novef

Dezenas

dezf; onzef; dozef; trezef; catorzef; quinzef; dezasseisf; dezassetef; dezoitof; dezanovef; eumf; edoisf; etresf'; equatrof; ecincof; eseisf; esetef; eoitof; enovef; vintef; trintaf; quarentaf; cinquentaf; sessentaf; setentaf; oitentaf; noventaf; vintee; trintae; quarentae; cinquentae; sessentae; setentae; oitentae; noventae

Centenas

edezf; eonzef; edozef; etrezef; ecatorzef; equinzef; edezasseisf; edezassetef; edezoitof; edezanovef; evintef; etrintaf; equarentaf; ecinquentaf; esessentaf; esetentaf; eoitentaf; enoventaf; evintee; etrintae; equarentae; ecinquentae; esessentae; esetentae; eoitentae; enoventae; cemf, duzentosf; trezentosf; quatrocentosf; quinhentosf; seiscentosf; setecentosf; oitocentosf; novecentosf; centoe; duzentose; trezentose; quatrocentose; quinhentose; seiscentose; setecentose; oitocentose; novecentose; ecentoe; eduzentose; etrezentose; equatrocentose; equinhentose; eseiscentose; esetecentose; enovecentose

Milhares

zero; um; dois; três; quatro; cinco; seis; sete; oito; nove; mil; milf; mile; e

Dezenas de Milhares

dez; onze; doze; treze; catorze; quinze; dezasseis; dezassete; dezoito; dezanove; vinte; trinta; quarenta; cinquenta; sessenta; setenta; oitenta; noventa; eum; edois; etres; equatro; ecinco; eseis; esete; eoito; enove

Centenas de Milhares

edez; eonze; edoze; etreze; ecatorze; equinze; edezasseis; edezassete; edezoito; edezanove; cem; duzentos; trezentos; quatrocentos; quinhentos; seiscentos; setecentos; oitocentos; novecentos

Milhões / Dezenas de Milhões / Centenas de Milhões

milhao; milhaof; milhoes; milhoes

Telefone

doisp9; zerop852; ump852; doisp852; tresp852; quatrop852; cincop852; seisp852; setep852; oitop852; novep852; zerop74; ump74; doisp74; tresp74; quatrop74; cincop74; seisp74; setep74; oitop74; novep74; zerop63; ump63; doisp63; tresp63; quatrop63; cincop63; seisp63; setep63; oitop63; novep63; zerof; umf; doisf; tresf; quatrof; cincof; seisf; setef; oitof; novef

Telemóvel

novep9; zerop6; ump6; doisp6; tresp6; quatrop6; cincop6; seisp6; setep6; oitop6; novep6; zerop358; ump358; doisp358; tresp358; quatrop358; cincop358; seisp358; setep358; oitop358; novep358; zerop247; ump247; doisp247; tresp247; quatrop247; cincop247; seisp247; setep247; oitop247; novep247; zerof; umf; doisf; tresf; quatrof; cincof; seisf; setef; oitof; novef

Número de identificação da segurança social

zerop4710; ump4710; doisp4710; tresp4710; quatrop4710; cincop4710; seisp4710; setep4710; oitop4710; novep4710; zerop369; ump369; doisp369; tresp369; quatrop369; cincop369; seisp369; setep369; oitop369; novep369; zerop258; ump258; doisp258; tresp258; quatrop258; cincop258; seisp258; setep258; oitop258; novep258; zerof; umf; doisf; tresf; quatrof; cincof; seisf; setef; oitof; novef

Data

do, de

Anexo 2: Lista dos sons gravados pelo locutor

Unidades

um; dois; três; quatro; cinco; seis; sete; oito; nove

(em posição final)

(contar até 10 mentalmente entre a leitura de cada numero)

Dezenas

dez; onze; doze; treze; catorze; quinze; dezasseis; dezassete; dezoito; dezanove; vinte; vinte e um; trinta e dois; quarenta e três; cinquenta e quatro; sessenta e cinco; setenta e seis; oitenta e sete; noventa e oito; setenta e nove; vinte e dois; trinta e três; quarenta e quatro; cinquenta e cinco; sessenta e seis; setenta e sete; oitenta e oito; noventa e nove

(contar até 10 mentalmente entre a leitura de cada numero)

setecentos e vinte; oitocentos e trinta; novecentos e quarenta; cento e cinquenta; duzentos e sessenta; trezentos e setenta; quatrocentos e oitenta; quinhentos e noventa

Centenas

nove mil e cem; oito mil e duzentos; sete mil e trezentos; seis mil e quatrocentos; cinco mil e quinhentos; quatro mil e seiscentos; três mil e setecentos; dois mil e oitocentos; cinco mil e novecentos;

nove mil cento e vinte; nove mil duzentos e trinta; nove mil trezentos e quarenta; nove mil quatrocentos e cinquenta; nove mil quinhentos e sessenta; nove mil seiscentos e setenta; nove mil setecentos e oitenta; nove mil oitocentos e noventa; nove mil novecentos e dez

(contar até 10 mentalmente entre a leitura de cada numero)

trezentos e dez; quatrocentos e onze; quinhentos e doze; seiscentos e treze; setecentos e catorze; oitocentos e quinze; novecentos e dezasseis; cento e dezassete; duzentos e dezoito; trezentos e dezanove; quatrocentos e vinte; quinhentos e trinta; seiscentos e

quarenta; setecentos e cinquenta; oitocentos e sessenta; novecentos e setenta; cento e oitenta; duzentos e noventa

trezentos e vinte e seis; quatrocentos e trinta e sete; quinhentos e quarenta e oito; seiscentos e cinquenta e nove; setecentos e sessenta e um; oitocentos e setenta e dois; novecentos e oitenta e três; duzentos e noventa e quatro

nove milhões e cento e vinte; um milhão e duzentos e trinta; dois milhões e trezentos e quarenta; três milhões e quatrocentos e cinquenta; quatro milhões e quinhentos e sessenta; cinco milhões e seiscentos e setenta; seis milhões e setecentos e oitenta; sete milhões e oitocentos e noventa; oito milhões e novecentos e dez

Milhares

mil; dois mil; três mil; quatro mil; cinco mil; seis mil; sete mil; oito mil; nove mil; três mil e vinte; quatro mil e trinta; e

Dezenas de Milhares

dez mil; onze mil; doze mil; treze mil; catorze mil; quinze mil; dezasseis mil; dezassete mil; dezoito mil; dezanove mil; vinte mil; trinta mil; quarenta mil; cinquenta mil; sessenta mil; setenta mil; oitenta mil; noventa mil

vinte e um mil; vinte e dois mil; vinte e três mil; vinte e quatro mil; vinte e cinco mil; vinte e seis mil; vinte e sete mil; vinte e oito mil; vinte e nove mil

Centenas de Milhares

cento e dez mil; cento e onze mil; cento e doze mil; cento e treze mil; cento e catorze mil; cento e quinze mil; cento e dezasseis mil; cento e dezassete mil; cento e dezoito mil; cento e dezanove mil

cem mil; duzentos mil; trezentos mil; quatrocentos mil; quinhentos mil; seiscentos mil; setecentos mil; oitocentos mil; novecentos mil

Milhões / Dezenas de Milhões / Centenas de Milhões

um milhão; um milhão cento e dez mil; dois milhões; dois milhões trezentos e quarenta mil

Telefone / Telemóvel / Número de identificação da segurança social

O número de telefone é

O número de telemóvel é

274 619 380

385 728 491

496 839 502

507 940 613

618 051 724

729 162 835

830 273 946

941 384 057

052 495 168

163 506 279

Data

A data é vinte do seis de 2014